RTO-TR-7
AC/323(HFM)TP/3

RTO-TR-7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO TECHNICAL REPORT 7

Alternative Control Technologies

(Technologies de contrôle non conventionnelles)

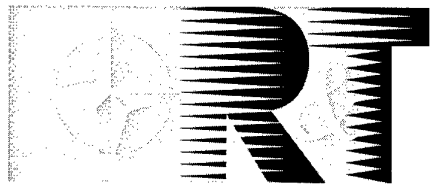
This Technical Report has been prepared at the request of the RTO Human Factors and Medicine Panel (HFM).

Published December 1998

Distribution and Availability on Back Cover

RTO-TR-7
AC/323(HFM)TP/3

NORTH ATLANTIC TREATY ORGANIZATION



RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO TECHNICAL REPORT 7

Alternative Control Technologies

(Technologies de contrôle non conventionnelles)

by

Dr Bernard HUDGINS, Canada
Dr Alain LEGER (Chairman), Dr Pierre DAUCHY, Mr Dominique PASTOR, France
Dr Hans PONGRATZ, Germany
Dr Graham ROOD, Mr Allan SOUTH, Dr Karen CARR, Dr Don JARRET, United Kingdom
Dr Grant MCMILLAN (Vice-Chairman), Dr Timothy ANDERSON, Mr Joshua BORAH, USA

This Technical Report has been prepared at the request of the RTO Human Factors and Medicine Panel (HFM).

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 6 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

Printed on recycled paper

Published December 1998

ISBN 92-837-1009-6

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Alternative Control Technologies

(RTO TR-7)

Executive Summary

In January 1996, the Working Group 25 of the former AGARD Aerospace Medical Panel began to evaluate the potential of alternative (new and emerging) control technologies for future aerospace systems. The present report summarises the findings of this group.

- Chapter 1 reviews the operational need for new modes of cockpit interaction. Included here are a discussion of the increasing input-output demands on pilots and the limitations of systems such as HOTAS (Hands-On Throttle and Stick).
- Chapter 2 is an extensive review of the state of the art in alternative control. It provides detailed discussions of the technology and system control applications of speech recognition, head tracking, eye-line-of-sight tracking, hand and facial gesture recognition, and electrical signals (biopotentials) generated by the muscles and brain.
- Regardless of the operational benefits afforded by a new control technology, this benefit will only be achieved with proper integration of the new technology into a cockpit system. Chapter 3 discusses the human factors tools that are available to support the integration process and enumerates the unique engineering constraints associated with the nonconventional control technologies.
- Based upon the operational need and the technological capabilities, Chapter 4 describes some near-term applications of alternative control. At present, speech recognition and helmet tracking are scheduled for incorporation in a number of fighter aircraft. Eye tracking appears to be the next addition to helmet-mounted systems in order to: (1) increase the effective aiming envelope, (2) ease the problem of head movement under acceleration, and (3) increase the speed and naturalness of the aiming process. Applications for the control of military wearable computers are also discussed in this chapter.
- Chapter 5 outlines needed enhancements to the technology and integration tools and summarizes the findings of the report in several tables.
- A glossary and five appendices are provided at the end of the report.

Some years ago, the HOTAS concept has brought about a very significant evolution at the cockpit level. Looking at current systems, we can foresee the limitations affecting these concepts by total reliance on manually activated controls. Alternative Control Technologies should alleviate difficulties found with current systems and offer new possibilities for further evolution. The global conclusion of this report is that these technologies should now be progressively introduced, based upon strict operational considerations and technical availability.

Technologies de contrôle non conventionnelles

(RTO TR-7)

Synthèse

En janvier 1996, le Groupe de travail 25 du Panel de Médecine Aérospatiale de l'ancien Groupe consultatif pour la recherche et les réalisations aérospatiales (AGARD) a commencé l'évaluation du potentiel de nouvelles modalités de commande pour les systèmes aérospatiaux futurs. Ce rapport présente les conclusions de ce groupe.

- Le chapitre 1 analyse du point de vue opérationnel les besoins en nouvelles modalités d'interaction dans le cockpit. Il discute notamment les exigences imposées aux pilotes en termes d'entrées/sorties ainsi que les limitations induites par des principes comme HOTAS (mains sur manche et manette).
- Le chapitre 2 constitue un état de l'art approfondi des nouvelles modalités de commande. Il fournit des critiques détaillées des technologies suivantes et envisage leurs applications possibles: reconnaissance de la parole, suivi de tête, détection de la direction du regard, reconnaissance de gestes manuels et faciaux, exploitation des signaux électriques produits par les muscles et le cerveau (biopotentiels). Quels que soient les gains opérationnels rendus possibles par de nouvelles modalités de commande, ils ne seront effectifs que moyennant une intégration adéquate de ces nouvelles modalités dans le cockpit.
- Le chapitre 3 étudie les outils d'aide à la conception ergonomique disponibles et énumère les contraintes particulières d'ingénierie associées aux technologies de contrôle non-conventionnelles.
- En se fondant sur les besoins opérationnels et les capacités technologiques mises en évidence, le chapitre 4 décrit quelques applications réalisables à court terme. Il est d'ores et déjà prévu d'intégrer reconnaissance de la parole et suivi de casque dans un certain nombre d'avions de combat. Le suivi du regard devrait constituer le prochain ajout aux casques afin : (1) d'augmenter le champ de visée efficace, (2) de limiter les problèmes liés aux mouvements de la tête sous accélération, et (3) d'accélérer et de rendre plus naturelle la visée. Ce chapitre étudie aussi l'application de ces nouvelles modalités de commande aux ordinateurs intégrés aux tenues ("wearable computers") dans un contexte militaire.
- Le chapitre 5 s'intéresse aux améliorations à apporter à ces technologies et aux outils d'intégration et récapitule les conclusions sous forme de tableaux. Le rapport s'achève par un glossaire et cinq appendices.

Il y a quelques années, le concept HOTAS a amené une évolution très significative au niveau du poste d'équipage. Le regard qui peut être maintenant porté sur les réalisations actuelles permet de prévoir les limitations apportées par l'utilisation de contrôles purement activés manuellement. Les technologies de contrôles non-conventionnelles devraient contribuer à l'amélioration des difficultés rencontrées dans les systèmes actuels el offrir de nouvelles possibilités pour des évolutions ultérieures. D'une manière générale, ce rapport permet de conclure que ces technologies devraient maintenant être progressivement introduites dans les postes d'équipages, en suivant des considérations strictement basées sur le besoin opérationnel et la disponibilité des techniques.

Contents

Preface

Working Group 25 was formed under the former AGARD Aerospace Medical Panel in January 1996 to address the following issue: Evaluate the state of the art and formulate a framework for the integration, evaluation and application of a variety of emerging control technologies in aerospace systems.

Six meetings were held over the next two years in order to accomplish this objective. At each meeting our host provided excellent working environments, highly relevant technical demonstrations and sincere hospitality. Each working group member would like to express their appreciation to the following hosts for supporting our work:

- Armstrong Laboratory, Wright-Patterson Air Force Base OH, USA, January 1996
- SEXTANT Avionique, Bordeaux, France, June 1996
- Kastellet, Copenhagen, Denmark, October 1996
- Flugmedizinisches Institut der Luftwaffe, Fürstenfeldbruck, Germany, April 1997
- Defense Evaluation and Research Agency, Farnborough, UK, September 1997
- Applied Science Laboratories, Bedford, MA, USA, February 1998

The present report is a clear example of the international cooperation that the former AGARD has successfully generated throughout its existence. The members are confident that it represents a significant contribution to the field of alternative control technologies and that it will support the application of these technologies in future aerospace systems. Each member contributed in many important ways to the success of Working Group 25 and we anticipate continued collaboration, long after the work of this group is completed. All participants feel privileged to have been associated with this work and such an imaginative and sincere group of scientists and engineers.

Préface

Le Groupe de travail 25 du Panel de Médecine Aérospatiale de l'ancien Groupe consultatif pour la recherche et les réalisations aérospatiales (AGARD) a été constitué en janvier 1996 afin d'évaluer l'état de l'art et d'élaborer un cadre de travail pour l'intégration, l'évaluation et l'application de différentes modalités de commande émergentes pour les systèmes aérospatiaux.

À cette fin, il s'est réuni six fois au cours des deux dernières années. À chaque réunion, les organismes hôtes ont fourni d'excellents environnements de travail, des présentations techniques pertinentes et une sincère hospitalité. Tous les membres du groupe souhaitent exprimer leur gratitude aux organismes suivants pour leur soutien à notre travail :

- Armstrong Laboratory, Wright-Patterson Air Force Base OH, États-Unis, janvier 1996
- SEXTANT Avionique, Bordeaux, France, juin 1996
- Kastellet, Copenhagen, Danemark, octobre 1996
- Flugmedizinisches Institut der Luftwaffe, Fürstenfeldbruck, Allemagne, avril 1997
- Defense Evaluation and Research Agency, Farnborough, Royaume-Uni, septembre 1997
- Applied Science Laboratories, Bedford, MA, États-Unis, février 1998

Le présent rapport est un bon exemple de la coopération internationale que l'ancien AGARD a suscitée avec succès tout au long de son existence. Les membres du groupe sont persuadés qu'il apporte une contribution significative au domaine des nouvelles modalités de commande et aidera à utiliser ces technologies dans des systèmes aérospatiaux futurs. Chacun des membres a apporté de multiples et importantes contributions au succès du Groupe de travail 25 et nous espérons continuer cette collaboration longtemps après l'achèvement du travail du groupe. Tous les participants sont honorés d'avoir été associés à ce travail, dans un groupe de scientifiques et d'ingénieurs aussi imaginatif et sincère.

Foreword

Man inherited a superior ability to interact with the environment using his hands. The major evolutionary step taken when this capacity was extended to manufacture stone tools, to enhance the direct mechanical action of the hand on "dumb" mineral, vegetable and animal elements in the environment, perhaps qualified him as "habilis" long before he became "sapiens".

Through our social needs, which we had in common with all animals, we could also influence one another's behaviour without direct contact, using posture, sound, facial expression and, for instance like dominant wolves, cause submission by "gaze fighting". However, it was the human acquisition of articulated speech, with its rich mixture of semantic, prosodic and affective cues which introduced new dimensions to remote communication, albeit with the co-operation of a receptive "intelligent agent". We should note that this use of the voice has been vital for military purposes from antiquity to modern times to control the movement of troops in battle, and coordinate their actions.

In the aeronautical field, following the first powered lift-off of the "avion" at the end of the 19th century by Clement Ader, the first controlled flight manoeuvre, a stable turn, was performed in 1905 by the Wright brothers. Here the only intelligent agent was the pilot himself, and since all else was wood, metal, rubber and fabric, it required the pilot to exert mechanical control actions, largely through his hands, However nowadays most modern aircraft are controlled directly by complex computational systems, and such mediation theoretically overcomes the need to hold, push, pull or twist a lever in order to effect control. This transformation is worth emphasising; because a machine controlled by computers looks like a computer to its operator, the interface between the man and the machine can be as flexible as that between a man and a computer. We therefore have the opportunity to evolve the pilot from "habilis" status to "sapiens".

The computers have themselves evolved. Those from the 60's and 70's had very limited "intelligence" and memory, which greatly exercised both the programmer's skill in optimising the programmes and the operator's skill in using intelligent strategies to compensate for his own and the machine's shortcomings. However, the situation on the machine side is now completely different and, although the machine still lacks "intelligence", it possesses a huge memory capacity and an ability to process data with extraordinary speed which is quite the opposite of human abilities, an imbalance which is most evident at their interface. Here work seems to be required in two areas: the direct interfacial mechanisms and the overall system design. The first is necessary to improve the physical modes of interaction. The latter is needed to make the machine more like a human so that it can accept high level instructions and, perhaps, eventually be capable of understanding the intentions and needs of the operator. There is a nice analogy here; the progress from programming a computer using machine-intelligible codes to the modern use of high-level languages is paralleled by the idea of the operator interacting with machines at present using numerous detailed machine-compliant actions and the future possibility of control via high level human-oriented intentions.

The challenge to engineers and cognitive scientists is obvious. As well as effort directed toward making information from the machine easier for the operator to perceive and understand, there is a need for "human-centred" control concepts. An excellent statement of the motivation was put forward by Rasmussen and Vincente as the "ecological interface" which, they suggested, should be constructed in such a way that it did not constrain the operator to work at a higher level of control than required by the situation. This concept may now have started with the adoption of novel head tracking and speech recognition systems in new generation aircraft like the Eurofighter, Rafale and the JSF, in which for instance, the pilot may give a short command to an "intelligent" speech recogniser rather than have to complete a lengthy sequence of alphanumeric key pressing actions.

It must however be borne in mind that the case for maintaining a progressive improvement to man-machine interaction, no matter how strong the theoretical argument, must be justified practically. The most evident case is that manual controls and switches, particularly those mounted in the grip-tops and referred to as "hands-on-throttle-and-stick" (HOTAS), are already too numerous. There is an obvious chance of erroneous selection. One difficulty, which is quite likely to be a major problem, is somewhat paradoxical in that it stems from human adaptability, and here there are two issues. Firstly, the human ability to do without something which he has never had may make the provision of it unjustifiable to a hard-nosed accountant. Secondly, our intrinsic ability to devise a suitable strategy for overcoming equipment limitations makes it difficult to predict how something can be used most effectively, and therefore determine how best to set it up. Perhaps, to exploit novel technologies the system integrator should build-in the flexibility to allow the user to adopt a strategy which best enables him to fulfil his objectives. In any event, the rationale for providing alternative controls must be made primarily in terms of a reduction in the effort, both cognitive and sensorimotor, which the operator expends in performing his job. This, and a substantially reduced chance of error, would be apparent in war as improved mission effectiveness and in peace as enhanced safety. For the provider, the cost of equipment supply and maintenance are likely to be countered best by quantifying the benefits in terms of a reduction in training time, although the gain in confident proficiency and general well-being of the operators should not be neglected. If built-in flexibility affects either the benefits to performance or the training needs, this must be included in the trade-off.

For two years Working Group 25 has tried to review comprehensively the issues associated with the implementation of Alternative Control Technology in the aerospace environment. Most of the information collected can be applied to other defence or civil applications, particularly the reviews of the states of the art in each of the technological areas. The issues which must be addressed in order to integrate these systems into the man-machine interface have been approached from both engineering and human factors viewpoints, and the need for further research has been identified, mainly as a set of challenges in the context of combat aircraft. It is recognised that a considerable amount of work remains, because very little will result from merely putting electronic boxes side by side and connecting them to the rest of the system.

Like the successful exploitation of automation, achieving meaningful and effectively integrated solutions will require the synergistic effort of a wide range of skilled individuals in research laboratories, equipment manufacturers and airframe manufacturers. We hope that our efforts will provide practical help in this endeavour.

Human Factors and Medicine Panel Officers

Chairman:

Dr. M. WALKER
Director, Centre for Human Sciences
DERA - F138 Building - Room 204
Farnborough, Hants GU14 0LX
United Kingdom

Membership of former AGARD AMP Working Group 25

Chairman

Dr Alain LEGER
Sextant Avionique
Rue Toussaint Catros
BP 91
33166 Saint Médard en Jalles Cedex, France

Vice-Chairman

Dr Grant MACMILLAN
Armstrong Laboratory
Human Engineering Div. (AL/CFHP)
Wright-Patterson AFB, OH 45433-7022, USA

Members

Dr Bernard HUDGINS
Institute of Biomedical Engineering
University of New Brunswick
Fredericton, New Brunswick E3B 5A3, Canada

Mr Dominique PASTOR
Sextant Avionique
Rue Toussaint Catros
BP 91
33166 Saint Médard en Jalles Cedex, France

Dr Graham ROOD
Manager Mission Management Dept.
DERA - Room 204 - Bldg R177
Farnborough, Hants GU14 6TD, U.K.

Dr Karen CARR
Sowerby Research Center
British Aerospace - PFC 267
P O Box 5, Filton
Bristol, BS12 7QW, U.K.

Dr Timothy ANDERSON
Armstrong Laboratory
Biodynamics & Biocommunications Div.
(AL/CFBA)
2610 7th Street, Building 441
Wright-Patterson AFB, OH 45433-7901, USA

Mr Pierre DAUCHY
IMASSA-CERMA
Institut de Medecine Aérospatiale
BP 73
91223 Brétigny sur Orge Cedex, France

Dr Hans PONGRATZ
Leiter ABT III
Flugmedizinishe Institut der Luftwaffe
P O Box 1264 KFL
82242 Fuerstenfeldbruck, Germany

Mr Alan SOUTH
DERA - Room 132 - Bldg R177
Farnborough, Hants GU14 6TD, U.K.

Dr Don JARRETT
DERA - Room 129 - Bldg R177
Farnborough, Hants GU14 6TD, U.K.

Mr Joshua BORAH
Applied Sciences Laboratories
175 Middlesex Turnpike
Bedford, MA 01730, USA

Panel Executive

Dr C. WIENTJES
RTA/HFM
BP 25
7, Rue Ancelle
F-92201 Neuilly-sur-Seine Cedex
France

1. OPERATIONAL NEEDS AND OPPORTUNITIES FOR ALTERNATIVE CONTROLS

1.1 INTRODUCTION

Combat aircraft can, in general, be described as manoeuverable airborne weapons platforms which contain a series of electronic and other systems with which the aircraft is controlled, navigated, weapons selected, etc., and a series of systems which provide protection for the aircrew throughout the performance envelope of the aircraft and when emergency escape is unavoidable. Most aircraft platforms have an operational life of over 20 years - some a lot longer - and, in this timescale, although the basic platform does not significantly alter - mainly for cost reasons - the avionics and crew support systems fits can continue to advance a number of generations - which can allow the airframe to retain its operational competitiveness against newer designs.

The rapid development of avionic systems in terms of reliability, speed and capacity, and the associated development of software, allows significant in-life updates to aircraft to be made at a more acceptable cost. The move towards Integrated Modular Avionics will allow systems to become more complex, with more data processing capacity at higher speeds, and with the amount of information output heavily increased. This is often all fed to a single pilot who is flying the aircraft close to the ground at around 450 knots or more, perhaps in bad weather at night, and the flying process alone needs continuous monitoring. In addition s/he needs to keep safe control of the aircraft, find the target, select and arm weapons, be aware of, and react to, enemy countermeasures, perform complex operations with smart weapons, etc., all in a degraded environment with high noise levels, high vibration and heat, high G levels, high agility, disorientation, etc. Out of this scenario, one of the primary problems is the amount of data - not necessarily in the right information format for easy digestion - that it is necessary for the pilot to process and the interaction with the displays which s/he will need to ensure that the correct inputs are entered at the right time, and quickly enough, to get the operationally relevant information out.

The more complex the new systems, and this increasing complexity is often needed to counter the increasing subtlety of enemy countermeasures, there is a tendency to need more inputs to a greater number of systems by the pilot and the additional time to carry out these extra operations is not generally available.

The current, and traditional, methods of data input or selection of systems normally require the use of the hands to either switch a system to a particular state or enter data through a keyboard. Most current aircraft, both civil and military, make large use of keyboards to enter a wide range of data both on the ground and whilst airborne. Errors do occur in data entry, even under benign conditions, and sometimes can result in serious consequences. In military aircraft, data entry is often an operational requirement in flight and experiments have shown that errors of around 2.2% to 2.9% can occur in high-speed low-level flight [1-1] and, even in the office environment, typing errors in the region of 1.5% occur, and this is with a full sized keyboard under unstressed conditions and without the need for nuclear, biological and chemical (NBC) protective gloves and the smaller keyboards and key sizes often found in aircraft. Key size differences can occur between a commercial keyboard and a military airborne keyboard - and there are recommended spacings in the human factors specification MIL -1472D. Next generation systems may need a larger number of data inputs and to increase the manual input capability of the pilot either requires an increase in 'typing' speed, a larger number of hands or an alternative control technique.

In civil systems errors occur most often during high workload periods [1.2] - often during a runway change required by Air Traffic Control during approach and, for the military, similar errors could be expected to occur in aircraft which use a combination of military and civil systems in the cockpit (C-130J, E-3D, C-17, etc), particularly, perhaps, in the more demanding battlefield support role.

More demanding operations in the current generations of fixed and rotary wing aircraft, particularly at night and in poor weather, have increased the need for more 'eyes-out' operations, which decreases the time for 'head down' or 'head in' viewing time, both for switching operations and for assimilation of information from head down displays. Similarly the speed of operations has led to less time being available for these two operations. Progress has been made towards the assimilation of visual display data through helmet mounted displays and the time reductions in switching have been achieved through ensuring that the pilot has no need to move his hands from the primary aircraft controls during high workload periods by the use of the **H**ands **O**n **T**hrottle **A**nd **S**tick (**HOTAS**) concept (or **H**ands **O**n **C**ollective **A**nd **C**yclic (**HOCAC**) for helicopters). Using Fitts' Law, namely that the time to move the hand to a target (in this case a switch or button) depends only upon the relative precision required, indicates that the movement time - a summed combination of perceptual processing, cognitive processing and motor processing - is in the region of 250 ms (an aircraft moving at 500 knots travels in the region of 80 metres in this time). Thus a time saving of around 250msec is achievable by minimising the hand movements. This generally involves the provision of all of the necessary manual switches on either the throttle top or the control column (stick) top during all critical flight operations. An example of HOTAS controls is shown in Fig. 1-1 for the AFTI F-16 aircraft [1-3.]

As the capabilities of aircraft will continue to increase through the use of more sophisticated, and a wider range of, sensors, and control through software increases, the ability to control the aircraft systems will inevitably require an even greater number of controls - many of these being necessary, at least in principle, on the HOTAS controls, as many are time critical and need to be operated eyes-out. The rise in the number of avionic systems and the consequent number of manual switching operations necessary during critical phases of operations (e.g. beyond the front edge of battle area (FEBA) and set-up and attack phase of a ground target) has resulted in a gradual increase in the numbers of switches/controls per crew member in the cockpit and this is illustrated in Figure 1-2.

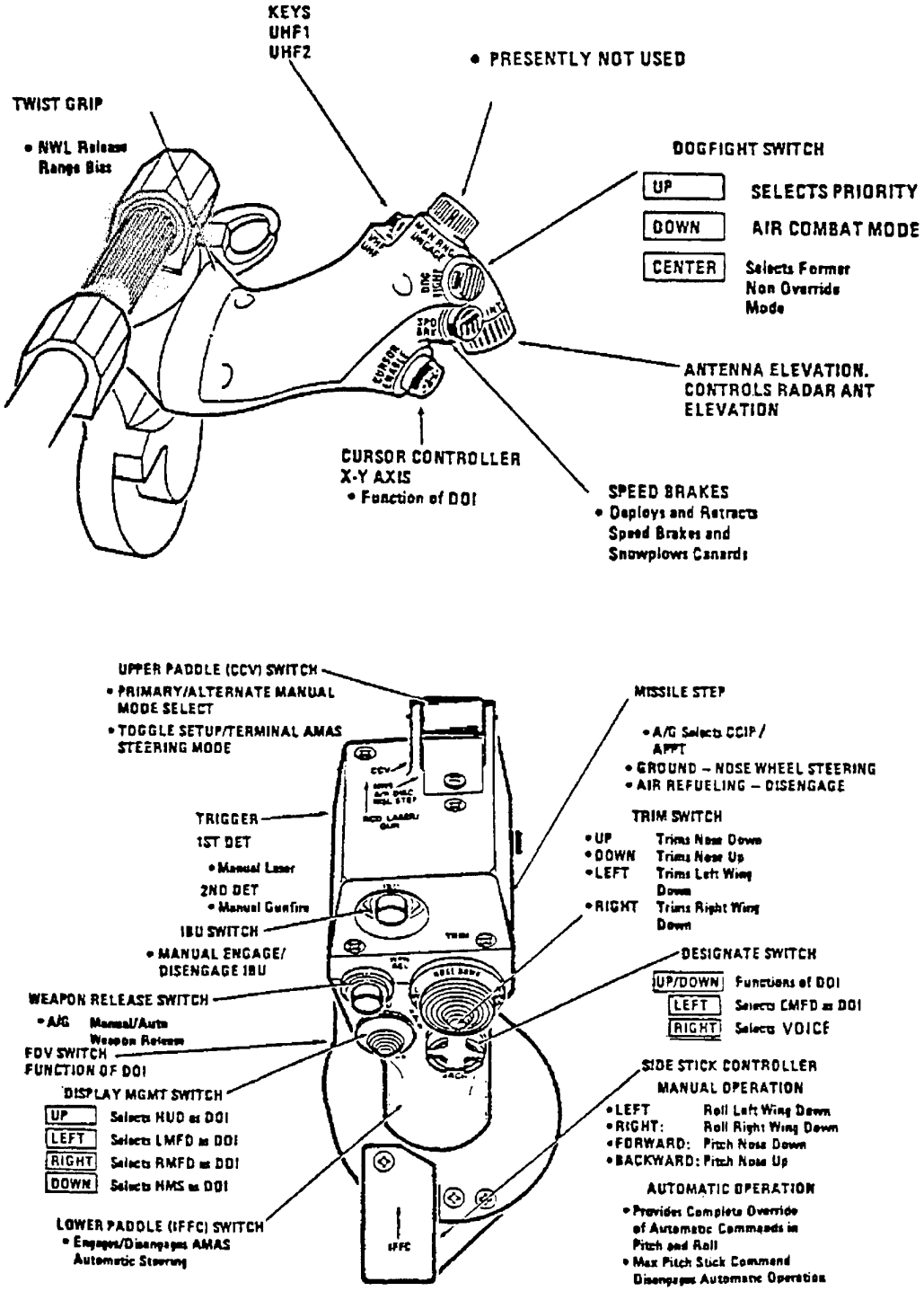


Figure 1-1 HOTAS Controls for AFTI/F-16 Aircraft

The increased numbers of switches and controls results both in longer selection and switching times and with the necessity to look head down into the cockpit to operate the correct switch or series of switches. This has led to the HOTAS concept and, on HOTAS, aircraft of the 1970's design era were using around 16 stick and throttle top functions, and, whilst some aircraft designs in the late 80's still used less than 20 functions, some fixed wing aircraft were up to 33 functions and helicopters up to 40. Figure 1-3 illustrates this trend and Table 1-I shows the functions allocated to HOTAS for a number of aircraft [1-3].

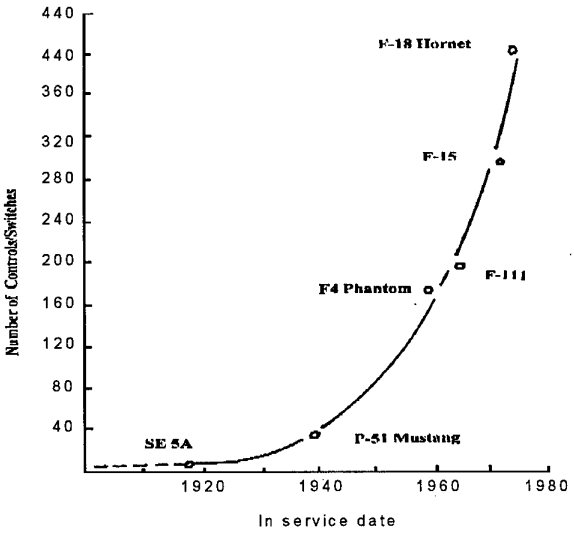


Figure 1-2 Number of controls and switches per crew member

There are some indications from aircrew that the numbers of functions are becoming both difficult to remember - needing more training - and sometimes difficult to operate with either standard aircrew gloves or NBC gloves. More complex systems will almost inevitably require more control mechanisms, and the most obvious approach is to increase the number of HOTAS keys - at least for the time critical operations. If the physical space is no longer available on the throttle or stick, the temptation will be to use 'chording' - the simultaneous use of two, or more, (existing) keys to select or operate systems - with an inevitable increase in mental complexity.

Since it is impracticable to label the switches - it would, in any case, be almost impossible to read the labels in their position in the aircraft or have the time to do so during critical parts of the sortie, - there is no possibility of identifying the correct switch or

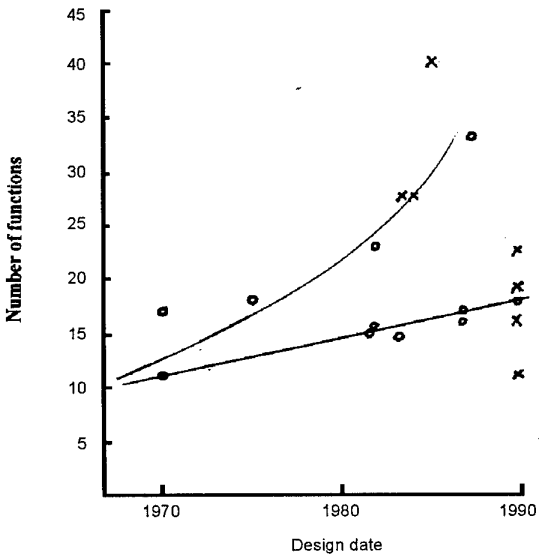


Figure 1-3 Trend of HOTAS Switching

Table 1-I A sample of functions allocated to HOTAS controls

Aircraft	Design Date	Throttle Functions	Stick Functions	Hand Controller	Total
F15C Eagle	1970	11	6	0	17
F15E-front	1982	9	6	0	15
F15E- rear	1982	0	0	6	6
Tornado IDS - front	1970	4	7	4	15
Tornado IDS - rear	1970	0	0	5	5
F-18 A toD - front	1975	10	8	0	18
F-18 E/F - front	1990	10	8	0	18+
F-18 E/F	1990	0	0	6	6
AV8B+	1989	9	8	0	16
Harrier GR7	1989	17(8)	17(8)	0	16+
Mirage 2000-5	1987	14	9	0	23
Rafale	1988	21	11	0	33
EF2000	1991	10	10	?	20+
AMX	1982	6	9	0	15
F-16 C/D Falcon	1983	6	8	0	14
AFTI F-16		8	10	0	18
MIG-29	1987	7	12	0	19
AlphaJet	1985	4	6	0	10
Tiger -rear	1985	14	12		26
- front		14	12		26
AH 64 Longbow-rear	1990	6	13	0	19
-front		0	0	11	11
EH101	1984	19 (14)	21 (12)	0	40
RAH66 Comanche	1990	14	8	0	22
MV22 Osprey	1988	9	7	0	16
A330 Airbus	1990	0	3	0	

button if the memory fails or falters for any reason. Since a large percentage of the buttons/switches are for critical aircraft functions, and thus will be time critical, any delay or error can jeopardise the aircraft mission. Further, even if the error is known, the procedures to recover from such errors - if any - inevitably take time. It may not always be clear to a pilot that s/he has made an error, or that s/he has pressed the wrong switch or button. If a button is pressed and the expected consequences do not occur, a number of options appear in his/her mind:

- The switch or button may not have worked:
  - Solution - press again or harder?
- The feedback system - if any - may have failed
- The display or function may have failed
- The system may have failed - is there any feedback?
- It may be the wrong button - which one now?

All of these take time, which generally is in critically short supply in these phases of flight. A well implemented **alternative** control input method would provide alleviation of this type of operationally critical problem.

A potential further problem, particularly with the necessary physical positioning of a larger number of switches or buttons is

Table 1-II Hand Length Data

	Date	Sample	mm	sd	Range	Spread
Male						
UK Military	1982	300	191.30	9.71	169-224	55
Canadian Military	1974	565	191.90	8.78	170-212	42
German Air Force	1966	1006	189.10	8.70	168-210	40
British Army	1970-75	2000	193.00	10.30	159-219	60
US Army	1970	1482	192.00	8.70	172-214	42
US Army	1966	6682	190.30	9.60	169-214	45
US Air Force	1970	148	197.20	9.30	173-228	55
French Army	1973	793	189.00	9.00	174-205	5th-95th% range
UK Civilian	1981	300	191.00	8.30	165-219	54
mean values			(191.65)	8.27	(159-228)	48
Female						
UK Military	1982	187	176.10	8.07	159-197	38
US Army	1977	1331	174.40	9.00	155-196	
US Air Force	1970	211	179.30	8.60	157-205	43
UK Civilian	1980	92	177.50	10.10	161-194	5th-95th% range
UK Civilian	1981	200	174.20	7.20	152-195	43
mean values			(176.30)	(8.6)		42.5

the difference in anthropometric span of the hand and fingers. Not only are there differences in the populations of an individual country, but there are statistical and practical differences between countries - sometimes significant. Currently, a number of countries are accepting female aircrew for combat aircraft, and the differences in HOTAS systems designed for male aircrew may elicit problems for female crew with differing effective digit length and hand-reach anthropometry.

Table 1-II shows an example of the differences in hand length of a number of countries and of a number of trials. The average hand length for males is 191.65 mm with an average spread of 48 mm. Standard deviations (sd) are in the region of 9 mm, which, as an estimate, would allow a HOTAS mounted set of switches and buttons to be designed to be used by perhaps some 70% (>1 sd) of the pilot population without undue difficulty. The remaining 30% may need to make some sliding movements around the stick or throttle to accommodate the full range. The female average hand length, however, is an average of 176.3 mm with a spread of 42.5 mm and an sd of 8.6 mm. The difference in mean length is some 16 mm, which could provide some difficulty in design of HOTAS controls which must be operated by both genders.

Table 1-III supports this hypothesis with figures comparing, in considerably more detail, differences between UK male and female hand dimensions [1-4]. As an indication of the potential problems, the distance from the 'hand crease' - representing, in this case, the apex of the HOTAS grip - to the finger tips displays an average difference of 1.2 cm. If a wider range of male and female crews need to be accommodated, then this difference may be increased to over 3 to 4 cm. Similarly for span between the thumb and the individual digits, which gives an indication of the ability to operate a thumb switch and another with one of the other digits average differences of around 1.3 cm are apparent. Similarly the true digit lengths - the length of each finger - is shorter for females by around 5 mm and the curved hand length is shorter in females by some 1.65 cm. Many of these differences

Table 1-III Details of Hand Dimensions

		Male			Female		
		Mean	Sd	Range	Mean	Sd	Range
Finger number to hand crease							
Digit 2	Left	12.26	0.81	10.2-14.4	11.16	0.70	9.4-13.4
B-PQ	Right	12.21	0.79	10.2-14.9	11.17	0.69	9.3-13.1
Digit 3	Left	13.51	0.92	11.1-16.0	12.12	0.79	10.3-14.4
C-PQ	Right	13.40	0.87	11.3-16.5	12.09	0.78	10.3-14.6
Digit 4	Left	12.44	0.95	9.8-15.0	11.00	0.86	9.0-13.5
D-PQ	Right	12.31	0.90	10.0-14.9	10.97	0.85	9.2-13.6
Thumb	Left	6.02	0.50	4.7-7.6	5.50	0.44	4.3-6.9
AU	Right	6.11	0.48	4.9-7.8	5.62	0.43	4.2-7.0
Digit 5	Left	9.13	0.98	7.0-12.3	8.28	0.80	6.2-10.9
EW	Right	9.49	0.90	7.1-11.8	8.31	0.89	6.3-11.1

may be able to be accommodated by good design, but there must be a high probability that, in current designs, and in future designs where the increasing number of controls surfaces will perhaps result in physically smaller switches and buttons, the potential competition between switch numbers and available surface area, as numbers of switches or tactile controls compete with surface area, will play a more significant limitation.

1.2 HEAD POINTING

Currently, the majority of aircraft carrying out a missile attack on a ground or airborne target must point the nose of the aircraft towards the target in order to suitably align the enemy aircraft on the weapon aiming displays on the head-up display (HUD) to lock-on the weapon prior to firing. This is not only a time consuming approach, but may require the aircraft to perform tortuous manoeuvres in pursuit of the also manoeuvering target aircraft. Figure 1-4 illustrates the sustained and instantaneous manoeuvre capability that is currently required from an air-to-air combat fighter, in this case the F-16.

Unfortunately the human body, being developed over a few million years for a less stressful environment, does not respond well to these violent manoeuvres and technologically complex and ingenious methods of protecting the body must be employed. Currently airframe soft limits in the region of 9g are in use in current production and future aircraft and the protection of the crew to these levels is complex and cumbersome.

The emergence of the technology, over the last 15 years, to allow flight worthy Helmet Mounted Displays (HMD) [1-5, 1-6, 1-7] and the development of accurate flightworthy Head Pointing Tracker Systems (HPS) has allowed methods other than manually boresighting the aircraft, to be used to enhance weapon delivery techniques.

Future-current and next generation weapon systems, particularly air-to-air close combat engagements, will be able use an alternative form of control system that will integrate the HMD, the HPS and the missile seeker head, Figure 1-5.

This will enable the missile seeker to be driven by the head pointing system to look in the direction that the pilot's head is pointing, and, as the pilot sights the target aircraft in his helmet mounted sight, for the missile to lock-on and be fired at high off-boresight angles, without the necessity for violent manoeuvering

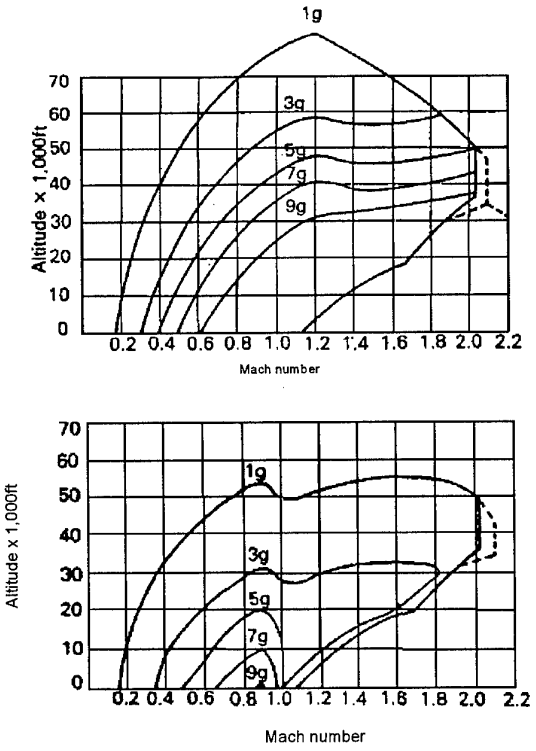


Figure 1-4 Instantaneous (upper) and sustained (lower) manoeuvre capability of the F-16/79

of the aircraft. Flight trials both in the USA, where live missiles have been fired at drones (BOXOFFICE) and in the UK, where air-to-air close combats have been carried out in 1 v 1 trials (JOBTAC) significant reductions in target acquisition and engagement times are apparent.

The use of a helmet mounted display and head tracking system in an F-16, combined with a missile capable of acquiring targets of over 60 degrees off-boresight, has allowed, in live firings against a QF 106 target drone at 0.7M, successful intercepts at 57 degrees off-boresight whilst the target was manoeuvering at 5g. Similarly, in one-on-one or two-on-two air-to-air combat between MIG-29s fitted with a simple Russian helmet mounted sight and using AA-11 (Archer) missiles, and F-16s with no helmet sight, the MIG-29 was able to attain the major number of first shot missile releases by use of the helmet sight system. To pass the head position information to the missile seeker, the MIG-29 used an electro-optical head tracking system. [1-8].

Similarly, at Farnborough in the UK, trials have been flown of one-on-one combat in a Jaguar, using a captive AIM9L and a standard Mk4 UK flying helmet fitted with a simple sight providing weapon systems information through a light-emitting diode (LED) display and an AC electro-magnetic head tracker. The system was developed by the Defence Research and Evaluation Agency (DERA) in collaboration with the General Electric Company Ltd. (GEC). Target acquisition and engagement times were significantly reduced, with off-boresight acquisitions up to 60 degrees being achieved.

As with most systems, however, whilst there may be significant operational shorter term advantages, there are also some longer term restrictions in the systems use of helmet mounted head

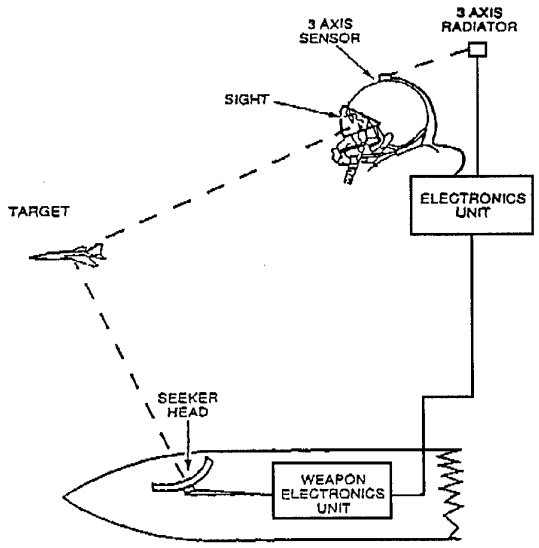
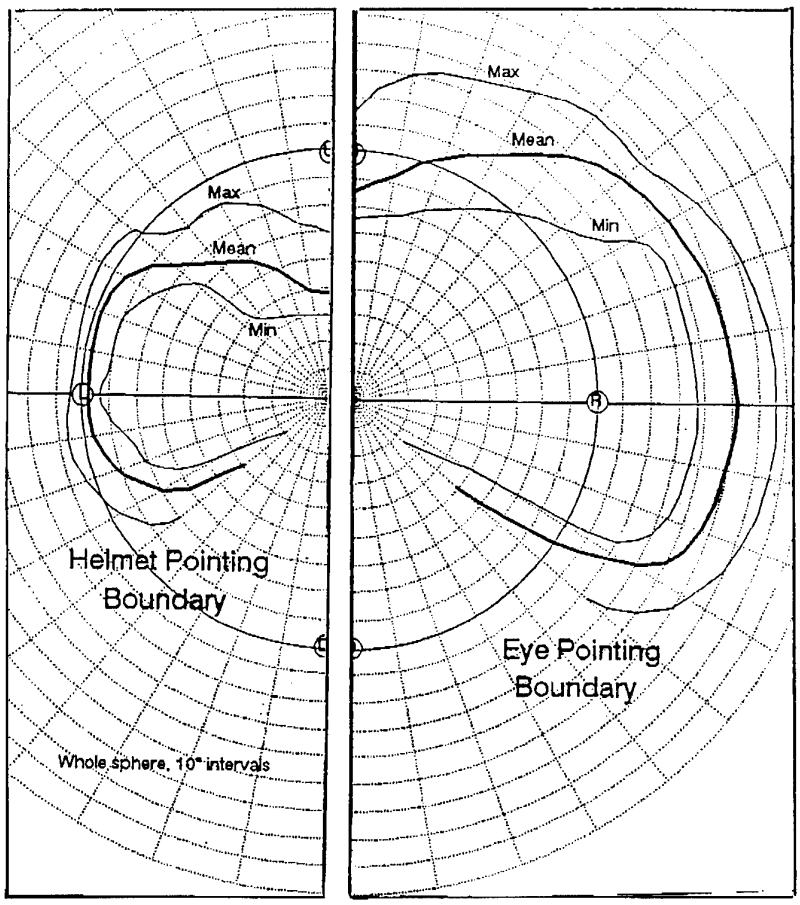


Figure 1-5 Use of head pointing in an off-bore-sight capability weapon system

pointing systems. One of those comes from the inability of a correctly strapped-in pilot to move his head much more than 90 degrees to the left or right. Figure 1-6 shows the head pointing envelope of a pilot, in full flying clothing, in a fast-jet strike aircraft cockpit, and, whilst the envelope is acceptable, it is limited by the available head movement of the human body. If, however, a further alternative control method, in the form of eye-tracking is utilised, then the useable envelope is significantly increased. This will allow, on average, tracking to around ± 140 degrees in the horizontal plane, compared to ± 90 degrees for head tracking and up to 90^0 in the vertical planes, compared to 55^0 with head tracking (in an aircraft with the restricted rearward and upward movement of the head from an ejection seat headbox).

Thus, it should be technologically possible to acquire targets in the rear hemisphere - or at least be able to input information into the weapon system as to the position of target aircraft outside the field-of-regard (FOR) of the conventional radar system or missile seeker [unless missile design changes] - but not, perhaps, outside of next generation thermal sensor's FOR. The Russian Vympel Design Bureau is reported as having tested a rear engagement capability in 1993 on a Sukhoi Su-27, the control authority of thrust vectoring allowing a rearward shot without the missile losing control as it initially flies backwards, [1-8].

Head tracking can also be used to designate ground targets from the air, or to point narrow FOV sensor systems at targets - and these generally replace manual control systems that are displayed on a head-down display (HDD). Hunting for a target, in a moving aeroplane, with a narrow FOV sensor (likened to looking for a target through a straw) can be difficult in the best of conditions and may take longer than is acceptable. By the use of either a helmet sight with head tracking, or with the addition of eye tracking, this type of operationally essential process can be considerably shortened and higher accuracies attained. UK trials have linked together such a system enabling the forward-looking infra-red (FLIR) sensor in TIALD (Thermal Imager and Laser Designator) to be located directly on a target of opportunity using a helmet sighted system in conjunction with the head tracker.



Comparison between helmet pointing and eye pointing envelopes

9 subjects
Mk10 ejection seat, tight harness
Tornado cockpit
Mk4 helmet, summer AEA and LSJ
(from RAE Tech Memo FS(F)-680 1987 by T K Manning)

Figure 1-6 Head and Eye tracking Envelopes

1.3 EYE TRACKING

Eye tracking has also some similar potential within the conventional cockpit or cabin, particularly with the use of large picture displays. These displays can either be in use in rotary or fixed wing strike aircraft, or in surveillance or command and control type aircraft. The problem lies in the use of a cursor in a large, and often cluttered, display, where the position of the cursor on the screen is not always immediately clear. For small FOV displays (say 20 deg x 20 deg) the cursor position can be determined more easily as it lies generally within the foveal cone of the eye and conventional manually controlled mice or joysticks are adequate. In a larger display, however, it can need considerably more scanning to find the cursor prior to repositioning it - with the obvious time delays. With conventional cursor control, it is necessary to find the existing position of the cursor in order to know which way to move the manual control to reposition the cursor at its new point By the use of eye tracking, however, it will be possible to reposition the cursor by the combination of fixing the eye on the required point and commanding the reposition with either a manual control or by the use of a voice command. This could also be used to reposition target boxes or similar designators in large screen displays, and combinations of eye tracking for coarse control and manual for fine control are feasible options. This combination of eye designation, manual fine control and target box labelling by voice command has the potential to provide significant reductions in aircrew workload.

1.4 VOICE CONTROL

Voice control or Direct Voice Input (DVI) has a large potential for alternative control techniques. In the HOTAS case, the problems may lie in the inability to remember either the position of the switch or the name of the function to be operated - more

probably the former than the latter. With the use of voice command to switch the system, the problem of memorising the switch or button positions is effectively nullified, and only the lesser problem of remembering the functions is left - in practice this should significantly reduce errors. Again, in practice, as with most **alternative** control technologies, it would be wise to retain redundancy in the system and allow operation by either manual and/or voice operated controls - pilot preference being allowed depending upon sortie patterns and phases. By using both systems, the number of manual operations on the HOTAS controls could be significantly reduced and HOTAS used for the time critical functions only, rather than its current potential for over-use - as there are no alternative control techniques to replace manual switching.

The use of voice control to select and switch systems has been discussed for a number of applications and is probably the lowest risk of alternative control technologies. Chapter 2.1 shows the development of the technology and it is clear that one major advantage over manual hard or soft key control is in being able to enter a, sometimes complex, hierachical control structure at any point. In most current systems (navigation, attack, etc.) it is necessary to page through the levels of a hierarchical menu to reach the level required. In the RAE (now DERA) Tornado flight trials, DVI was used on the navigator's TV-TABS and it was possible to access different levels of the navigation hierarchy directly with potential time savings. Whilst later systems have a less time consuming approach to the ability to access deeper parts of the system hierarchy, there remain structural problems with this approach, and whilst considerable ingenuity has been expended on reducing the number of button presses to access the required information, only manual keyboarding or voice control will allow direct access to the functions

Other areas that would benefit from the use of DVI are in the areas of radio channel selection. Currently, when a pilot needs to talk to a new controller, ground control, approach, tower, forward air controller, etc, it is necessary to obtain the frequency and select it on the appropriate radio - VHF/UHF/HF etc., before transmission. This process of obtaining the required controller, say Paris Orly approach, leads through mentally remembering the required frequency or looking up the frequency, through manual selection of the frequency on the appropriate radio and finally transmitting and talking to Orly approach - the person you first thought of - is unnecessarily time consuming and in many military operations time will matter. Voice command will shorten this process by asking, in a single operation, for Paris Orly Approach directly - the avionics will do the rest by recognising the request and having the frequencies already allocated to the controller in the avionics.

1.5 UNMANNED AIR VEHICLES (UAVS)

Over the next decade there is likely to be an increasing transition from air based cockpits to ground based cockpits for use with man-in-the-loop Unmanned Air Vehicles (UAVs). In the manned aircraft, the trend is likely to be, at least in a large number of air-to-ground operations, to isolate the human crew, as much as possible from the risks associated with combat areas. The natural trend, which is already visible from recent conflicts, is to produce stand-off weapons, either autonomous or with a man-in-the-loop control capability. Currently this is done from an airborne platform situated far enough from the target to minimise the risk of loss of, or damage to, the aircraft. As data links improve, by increased distance, immunity to jamming and increased bandwidths, the controlling site will be able to move to larger aircraft platforms and finally to ground borne stations. In each of these ground stations (ground or air based), control can be of either UAVs which are intended to fly returnable missions - or UAVs which are not intended to return to base.

Movement of the control station to the technically, and environmentally, more friendly ground station has a number of obvious advantages. Noise, vibration, heat, and those discomforts and partial disablers associated with aircraft manoeuvres - high G for example - are not present and the encumbrancies necessary for aircrew protection - laser protection, flying helmet, oxygen mask, G suit, NBC personal equipment etc., - are eliminated. Other factors, such as displays equipment, do not require the airborne equipments limitations on mass and volume to be implemented, nor do associated issues such as display brightness and display power. This should allow Commercial Off the Shelf (COTS) avionics equipment to be more utilised which will significantly support the affordability of these type of military operations.

Consequently, the use of alternative control technologies to supplement the natural human performance, often in terms of speed and accuracy, rather than compensate for the inadequacies and compromises that are essential in the cockpit environment, are more viable.

For instance, head-tracking systems are not exposed to unwanted motion from ground induced turbulence during ground attack sorties, voice system recognition rates improve in a low noise and vibration free environment, eye tracking devices will not require the complex integration into the airborne flying helmet and devices that are sensitive to environmental infra-red emissions (e.g. sunlight) can be more readily used - if appropriate.

The benefits of using alternative control technologies are not only apparent in the severe military air environment. The ability to operate more naturally with avionic and military systems, even in the more benign environments of the surveillance aircraft or the ground-borne UAV cockpit, should provide significant benefits to military operations.

1.6 CONCLUSIONS

Future manned cockpits will inevitably have more complex avionic fits to cope with more demanding operational scenarios and aircraft roles, and there will need to be an advance in the way that aircrew interface with the aircraft systems in order to enable efficient control between man and the rising complexity of aircraft systems. The number of manual control systems, including buttons, keyboards, and switches, is reaching a point where training aircrew to remember the phases and modes of switching could become both a significant proportion of operational training cost and also have flight safety implications. Similarly the increasing number of switches on HOTAS controls has the potential to heighten confusion rather than provide solutions. What is required are **alternative** methods of inputting data to aircraft avionic systems, particularly if they provide a more natural, and quicker, interface. A simple example of this is in the use of voice input as an alternative to remembering and dialling up radio frequencies. A single command phrase - Farnborough Tower - for instance, replaces, essentially, a three segment approach - remember frequency, dial frequency and call controller on that frequency. Of the more mature alternative control technologies, voice recognition and head tracking are both in operational flight and experimental flight - depending upon the level of sophistication of the technology - and are both technically

mature enough for full operational use, with research on the next generation, higher capability, systems in progress.

Eye based control is laboratory mature, and used for assessing eye movement in simulators, and, with development, has the potential to integrate effectively in the operational environment with head and voice based control. Gesture and biopotential are probably the least mature, but provide potential for the longer term aircraft systems (2020) and may be particularly of use in ground based cockpits of man-in-the-loop UAVs.

All systems in a civil and military aircraft must provide some tangible operational benefit - particularly in retrofit cases - and both head and voice based control are expected to provide that benefit in the third generation aircraft (Eurofighter and Rafale). This would be supplemented, in due course, with eye based control, particularly in the air-to-air engagement role, but, also, to a lesser extent, in the air-to-ground role.

The benefits of alternative control techniques lie in a more natural interface with the aircraft, improved speed of operation and reduction in training overheads.

Released from the constraint of only one communication channel with the aircraft systems - manual - the use of alternative control technology invites aircrew, aircraft and systems designers, and others, to be more imaginative in their interaction with the aircraft and systems, using these alternative controls as appropriate to the operational benefits and needs. Such alternatives are not intended primarily to replace manual controls but to supplement manual systems and to provide alternatives, to be used as the occasion requires. Aircraft systems, however, need to be practical, to retain as simple an interface as the technological complexity of the systems allows and be operated by aircrew with a wide range of capabilities. This should ensure that the use of these alternative controls is balanced by the aircraft designers natural, and often historically justified, inherent scepticism of the useability of new technologies.

1.7 REFERENCES

1-1 White G. and Becket, P. "Increased Aircraft Survivability using Direct Voice Input" AGARD CP 1983

1-2 FAA Human Factors Team Report on: The Interfaces between Flight Crews and Modern Flight Deck Systems June 1996

1-3 Flight Vehicle Integration Panel Working Group 21 Glass Cockpit Operational Effectiveness AGARD-AR-349, 1996

1-4 Gooderson C.Y.*et al* "The Hand Anthropometry of Male and Female Military Personnel" Army Personnel Research Establishment Memorandum 82M510, 1982

1-5 The Man Machine Interface in Tactical Aircraft Design and Combat Automation. AGARD Conference Proceedings No 425 (AGARD-CP-425), 1987

1-6 Combat Automation for Airborne Weapon Systems: Man Machine Interface Trends and Technologies AGARD Conference Proceedings 520 (AGARD-CP-520), 1993

1-7 Advanced Aircraft Interfaces: The Machine Side of the Man Machine Interface AGARD Conference Proceedings 521 (AGARD-CP-521), 1992

1-8 Aviation Week 16 October 1995

2. REVIEW OF ENABLING TECHNOLOGIES

This chapter reviews state of the art in speech-, head-, eye-, gesture- and biopotential- (muscle and brain electrical activity) based control. Each major section begins with a statement of the intention of the technology and why it is appropriate as a human-machine control modality. Next, various approaches for implementing the control modality are described, including a discussion of the strengths and weaknesses of each approach. This is followed by a discussion of current applications of the technology. Each major section concludes with a summary of needed enhancements and the prognosis for achieving them in the near future.

2.1. SPEECH-BASED CONTROL

2.1.1. INTENTION OF THE TECHNOLOGY

Current speech-based control systems are the most mature of those discussed in the chapter. Although research in this area goes back over 25 years [2.1-1], applications are only recently becoming widespread and accepted by the user community. This is due for the most part because of both limits in the technology and the very high expectations of the technology. It must be highly accurate, robust, and reliable to meet user needs and expectations. Speech-based control systems must be easy to use, that is, transparent to the user. The system should adapt to the user; not force the user to adapt to the system. In the following sections, a brief tutorial of terminology and components of speech-based control will be presented.

When discussing automatic speech recognition (ASR) systems, it is convenient to subdivide them into classes according to the problems they address. Systems are usually first divided according to the number of speakers they recognize.

Speaker-dependent systems can recognize speech from only one speaker, the speaker that trained the system. Speaker-independent systems recognize speech from many speakers, not only the speaker that trained the system.

The next subdivision that occurs for ASR systems is based on how they handle word boundaries. Isolated word recognition systems require a 100-250 ms or longer pause inserted between spoken words. Connected word recognition systems require a very short pause between words. Continuous speech recognition systems require no pause between words and accept fluent speech.

An additional subdivision that occurs for ASR systems is based on the size of the vocabulary or number of words that the system can recognize. Vocabulary size is usually divided into small (less than 200 words), large (1000 to 5000 words), very large (5000 words or greater) and unlimited (greater than 64000 words).

When defining a vocabulary for a specific task, a grammar may be developed that specifies which words may follow other words. This syntax, when incorporated into the recognition algorithm, has the effect of reducing the total number of words that must be considered by the recognizer at any one time. This improves both the speed and accuracy of the recognizer. Perplexity is a common metric used to determine the complexity of a grammar. Perplexity is defined as the average branching factor of the grammar or, stated another way, the average number of words that can follow each word in the grammar. The larger the perplexity of a grammar, the more difficult the recognition task.

Which combination of characteristics is best? The answer depends on the particular application that one is trying to accomplish with speech-based control and the characteristics of the user, task, and environment.

2.1.2. OVERVIEW OF APPROACHES

Speech generation is described by means of the "Source-Filter" model: a source of sound energy, which may be regular pulses from the vocal chords, or random fluctuations in the pressure of air being forced though a narrow constriction, is applied to a cavity with many resonant frequencies (i.e. the vocal tract). The frequencies and bandwidths of the resonances are determined primarily by the shape of the tongue, but also by the positions of the jaw, lips and velum.

In normal usage, speech carries several different kinds of information. As well as the semantic content, there is also information about the physical and emotional state of the speaker and cues to control the dialogue between speakers. The speech signal will be modified by the microphone and subsequent signal conditioning. In control applications of speech recognition, only the semantic content of the speech signal is required, so all the other kinds of information tend to act as perturbations which reduce the recognition performance. The speech signal could also be used to monitor the speaker's physical or emotional state (see section 2.1.3.3.).

Automatic speech recognition can be viewed as a pattern recognition task that maps an input speech waveform to its corresponding text. Although a wide variety of specific components and processes have been used, all speech recognition systems consist of combinations of the following functional elements:

- Signal acquisition -- microphones of various styles and frequency responses.
- Signal processing -- digital signal processing algorithms that identify or quantify the speech signal.
- Pattern Matching -- algorithms that transform the processed speech into a text string of the recognized speech.

Each of the components will be described in the following sections.

2.1.2.1. Signal Acquisition.

The speech signal is characterised by variation of energy with both time and frequency. The frequencies of interest lie between about 100 Hz and 8 kHz, although a narrower

bandwidth can suffice to carry intelligible speech. Ordinary telephones transmit frequencies from 300 Hz to 3400 Hz. In the time domain, the most rapid variations typically occur over durations of a few milliseconds. At the upper end, some vowel sounds, and other features, may remain relatively stable for 100-200 ms.

The most commonly used microphones are the close-talking head-set microphone and the telephone hand-set, although other possibilities are lavaliere, desk-top, and array microphones. Each different microphone presents it own unique challenges because of various frequency characteristics and signal strengths due to the microphone or the mode in which it is used (i.e., a desk-top or array microphone allows the user to walk around the room, resulting in various signal strengths as a function of the user's relative position to the microphone). These challenges are even greater for speaker-independent systems where different microphones were used for training than those used in the desired application.

In military aircraft cockpit applications, the microphone is included in an oxygen mask. The transfer function is then due to the influence of the microphone and the acoustic cavity. The resulting transfer function is widely imperfect and, even if it is sufficient for speech communications, it must be balanced (flat) for speech recognition. One way to solve the problem is to incorporate pre-emphasis filtering in the signal parameterization chain. The second solution is to use microphones of better quality and to design new oxygen masks, in order to provide a transfer function as flat as possible. This second solution is obviously more complex than the first one, and could be adverse to some constraints the oxygen mask must respect. For example, under over-pressure, the pilot's security and integrity remain more important than speech recognition. In the case of rotary wing applications, the same problem occurs as soon as oxygen masks are used; but in some rotary wing applications, the pilot uses a differential close-talking head-set microphone. Due to the environmental noise, a pilot puts the microphone as close as possible to his mouth. In this case, the acquired signal involves electronic saturation. Such a problem can be easily solved by training in the same conditions (without noise but with a microphone position analogous to real flight), or by adjusting the audio return so that the pilot positions the microphone further from his mouth. For instance, the MultiHelicare application described in section 2.1.3.1.2. involved such problems, which were solved as described previously without loss of performance.

Gain control is also a practical problem which can greatly effect speech recognition performance. Speech acquisition is provided by analog tools, but in order to compute speech features to be recognized, an analog-to-digital converter is required. This analog-to-digital converter involves a processing gain which must be adjusted in order to avoid overflow during numerical computations. But since speech is a highly varying signal, gain adjustment must be accurate. If the gain adjustment is not dynamic, some speech sounds will be coded over a very few bits, without using the dynamic range of the converter and introducing quantization noise. In order to optimize the quantization dynamic range, an automatic and adaptive gain control is required. One would think that classical Automatic Gain Control (AGC) methods are sufficient, but this is not the case: if the speech level is too variable, the AGC can be adverse to speech recognition.

In most systems analog-to-digital conversion is performed at sampling rates of 8000 Hz or higher. The speech power in specific frequency bands is then estimated with Fast Fourier Transforms, digital band-pass filters, or some auditory modeling techniques. These signal processing techniques will be discussed in the next section.

2.1.2.2. Signal Processing.

Before the pattern matching stage of speech recognition can take place, it is necessary to transform the speech waveform into a more tractable representation. This is necessary to reduce the quantity of data which the pattern matcher must handle. A second, but related, purpose is to extract those features of the speech signal which carry the information that discriminates between words, while eliminating features that carry other types of information. Information relating to the pitch of the signal is generally discarded for purposes of speech recognition (at least in European languages - pitch may be important in tonal languages such as Mandarin).

In most cases, digital speech processing is preceded by a discrete pre-emphasis filter which compensates for the natural decrease of 6dB/octave due to human speech production. A classical filter is given by the following formula:

$$H(z) = 1 - 0.95z^{-1}$$

whose transfer function is shown in Figure 2.1-1.

Although there are many different ways of representing the speech signal, most of them have certain features in common. Almost all techniques produce some kind of representation of the short-term power spectrum over a period of 5-30 ms.

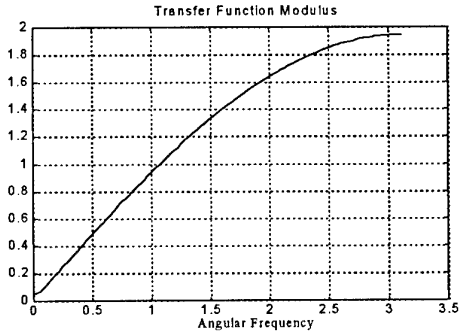


Figure 2.1-1 Pre-emphasis filter transfer function

Speech is a quasi-stationary signal; the spectrum may be approximately constant over periods of a few tens of milliseconds. It may also change rapidly within a few milliseconds, in plosive consonants, for instance. The purpose of windowing is to select a finite portion of the signal, which may be considered stationary, for analysis. The length of the window must be a compromise between spectral and temporal resolution. A long window will give a high resolution spectrum, but may hide the more rapid changes in the signal, whereas a short window will reveal the temporal structure more precisely, but blur the spectral characteristics.

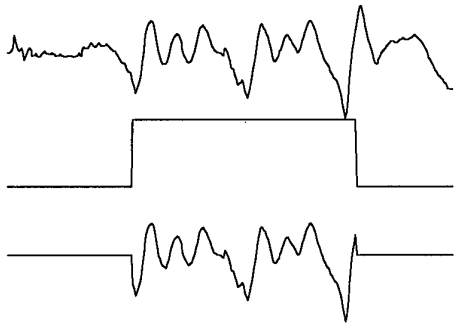


Figure 2.1-2 Rectangular window: Original signal (top); rectangular window (middle); windowed signal (bottom).

Window lengths of between 10 and 30 ms are commonly used for speech analysis

Mathematically, windowing is equivalent to multiplying the signal by a function which has value between 0 and 1 within the window and 0 at all other times. The simplest window is the uniform, or rectangular, window of length N samples:

$$w(n) = 1, \quad n = 0,1,....N-1$$
$$= 0, \quad \textit{all other } n$$

Figure 2.1-2 shows a frame of a signal extracted with a rectangular window. The temporal properties of the signal have been changed by this process, i. e. the new signal is zero outside the window. As a consequence, the spectrum of the signal is also inevitably changed. The spectrum of the windowed signal is obtained by convolving the spectrum of the original signal (assumed stationary) with the spectrum of the window [2.1-2]. The window spectrum is similar to a low-pass spectrum, with a broad main lobe at low frequencies and attenuated sidelobes at higher frequencies. The ideal window response will have a very narrow main lobe and large attenuation in the sidelobes. This can only be achieved by using a very long window, which defeats the object of using a window in the first place.

The rectangular window has a narrow main lobe for its length, but the attenuation in the sidelobes is very poor, only around 20 dB. A broad main lobe can be tolerated more easily than poor sidelobe attenuation, as the former reduces the local resolution while the latter spreads energy from distant parts of the spectrum. (In speech signals adjacent frequencies tend to be quite highly correlated anyway, so the local resolution is less important.) For this reason, many attempts have been made to design windows which reduce the sidelobes as much as possible. This is achieved by tapering the edges of the window in some way. Figure 2.1-3 shows the widely used Hamming window, which is described by the following:

$$w(n) = 0.54 - 0.46\cos\left(2\pi n/(N-1)\right), \quad n = 0,1,\ldots N-1$$
$$= 0, \quad \textit{all other } n$$

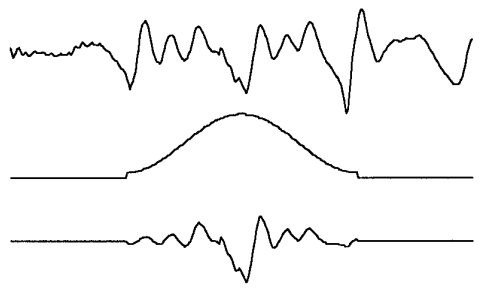


Figure 2.1-3 Hamming Window: Original signal (top); Hamming window (middle); windowed signal (bottom)

The sidelobe attenuation of the Hamming window is about 30 dB greater than that of the rectangular window. Note, however, that the samples towards the edges of the window are considerably attenuated, so it is important to overlap the windows for successive frames. If this were not done, important features of the signal which happened to fall on the boundaries between frames would not be given due prominence in the final signal representation.

Many other window designs are possible, although only a few are commonly used, such as von Hann, Hamming, Kaiser, and Blackman [2.1-3]. Which is best is dependent on the application, though the Hamming window is probably the most common in speech recognition front ends.

Following windowing, the frame is analysed by one of many possible methods, resulting in a string of about 10-20 numbers called a vector. In many cases, elements derived from the rates of change of the basic vector elements are added to the vector. The following paragraphs describe the most commonly used signal representations and discuss their various advantages and disadvantages.

The simplest representation of the speech signal is achieved by passing it through a parallel bank of bandpass filters. There are usually between 10 and 20 filters, covering the band from 200 Hz to 4 kHz. The bandwidth of each filter varies according to its center frequency, typically from 200 Hz at the low frequency end to 500 Hz at the high freqency end. The output of each filter is rectified and smoothed with a low-pass filter (cut-off usually about 25 Hz). The resulting value is sampled at the frame rate (50-100 Hz) and may be used directly or (more usually) compressed by taking its logarithm. An equivalent representation can be achieved by means of a Fourier transform followed by summation of the components within each frequency band.

An alternative way of representing the spectrum is to derive the parameters of an all-pole filter having the same response as the vocal tract at that point in time. This representation is known as Linear Prediction because of the technique used to calculate the filter coefficients using a linear combination of past waveform samples to predict the next sample. Many different methods exist to calculate these filter coefficients. See Rabiner and Juang [2.1-4] for a review of these different techniques and their advantages and disadvantages.

Several signal representations model, with varying degrees of accuracy, the processes believed to be used by the human auditory system. The motivation for this derives from the

fact that speech has evolved in conjunction with hearing and therefore, the nature of speech is heavily dependent on the capabilities of the ear.

Perceptual Linear Prediction (PLP) [2.1-5] implements three concepts from hearing to estimate the auditory spectrum: (1) the critical-band spectral resolution, (2) the equal-loudness curve, and (3) the intensity-loudness power law. The auditory spectrum is then approximated by an all-pole model (the same basic idea as Linear Prediction discussed above).

The filter bank described above may be regarded as a very low resolution auditory model. The main analogies with the human ear are that the bandwidth of the filters increases with frequency (the mel scale, [2.1-2]) and the amplitude response is logarithmic. At the other extreme, a full auditory model may have 100 channels and provide an output that mimics the firing of the nerves which carry signals from the ear to the brain. The computational power required for this kind of signal representation is very high.

The so-called cepstrum is derived by transforming the speech signal into the frequency domain with a Fourier transform, taking the logarithm of the power spectrum, and then using the inverse Fourier transform to return to the time domain. This gives a representation akin to a spectrum, but the horizontal axis is time (hence the name "cepstrum"). It is easy to distinguish between the pitch component and those components which represent the shape of the vocal tract.

Several of the basic signal representations may be greatly improved by subsequent processing using a technique known as Linear Discriminant Analysis. This is an optimization technique, applied during the development of the recognizer, or possibly during the training of the word models, which is used to find the combinations of channels and/or channel deltas which are best able to discriminate between the words of the vocabulary. The best known version of this technique is the IMELDA (Integrated mel scaled Linear Discriminant Analysis) transform [2.1-6]. The effect of applying the transform is to concentrate information into a small number of channels with little correlation between channels.

While a general transform can be derived for a given recognizer, this technique can be optimized for specialised applications, such as military aircraft. This gives a worthwhile improvement in performance.

The analysis of the human cochlea takes place on a nonlinear frequency scale, known as the Bark or mel scale. This scale is linear to about 1000 Hz and is approximately logarithmic above 1000 Hz. It is common to perform such a frequency warping for representations of speech. The most commonly used method of feature representation is that of mel-frequency cepstral coefficients or MFCCs [2.1-7]. MFCCs are generally computed every 10 ms by first performing a spectral analysis using a Fast Fourier Transformation on a window of 20 ms of speech. The spectrum is then warped using the above-mentioned mel-frequency warping. The logarithm of this warped spectrum is taken and followed by an inverse Fourier transform. The result is called the mel-cepstrum. By keeping the first dozen coefficients of the cepstrum, the spectral envelope information is preserved. The resulting features are the MFCCs.

The Fourier transform is one of the basic signal analysis tools relevant to analyzing stationary signals. But in the case of short-duration phenomena such as unvoiced plosives (/p/, /t/, /k/), the Fourier transform becomes less accurate. The wavelet transform (see Appendix A), which appeared in the last decade, has been introduced in order to process such non-stationary signals. Such decompositions may provide speech processing and acoustic pattern computation, which can be used by a pattern recognition algorithm. But, thanks to their mathematical foundations, these techniques can powerfully be used as speech feature extraction algorithms. Section 2.1.2.5. describes how such techniques are applied to acoustic phonetic decoding, error detection, and control.

A feature vector computed by one of the methods described above is used as the input to the pattern matching stage which is described in the next section.

Speech processing's influence on speech recognition performance is obvious. Classical speech processing can be improved by various algorithms (speech/noise discrimination, denoising algorithms, ...), whose aim is to take into account the particular environmental characteristics of military applications. These techniques will be described in section 2.1.4.1.

2.1.2.3. Pattern Matching

The pattern matching process consists of comparing the incoming speech with stored representations, which are usually whole-word models but may be phoneme-based. The word model that is most similar to the speech is considered to represent the word spoken. Both the incoming speech and the word models will be represented by sequences of vectors, so to achieve the comparison, one needs some means of measuring the similarity of the vectors and a way of determining which speech vector corresponds to which vector of a model.

The "distance metric" used to measure the similarity between vectors will depend on the signal representation used. The simplest is the Euclidean Distance, i.e. the sum of the squares of the differences between the individual components. Strictly speaking, this is only appropriate if all elements of the vector have the same significance, but factors are usually applied to give most weight to those channels known to carry most information.

In general, the correspondence in time between the vectors of the speech and those of the models is unknown. Even if the times at which a word starts and finishes are known (which is not usually the case), variations in the rate of speaking occur within words. Some speech sounds have relatively constant durations, while others vary widely. It is necessary, therefore, to find the optimum correspondence between the vectors of the incoming speech and those of each model. If the endpoints of the spoken word can be determined, it is possible to use linear time compression, but this is far from the optimum and is only practical for isolated word recognition.

Dynamic Time Warping (DTW) is the simplest means of optimizing the matching between vectors of the incoming speech and those of the models (see Appendix B). It is most often used in combination with simple models, such as stored sequences of vectors from single utterances of each word. A detailed description of the algorithm is given in Rabiner and Juang [2.1-4]. In outline, a distance score is calculated between each vector of the speech and each vector of the word model. It is then possible to find a sequence of vectors

from the model (some of which may be repeated and some may be skipped) which gives the minimum cumulative distance score. This is done using a mathematical technique called Dynamic Programming (or the Viterbi algorithm). The score for each model is normalized to allow for different numbers of vectors, then the model with the lowest score is taken to represent the word spoken.

For years, researchers have been developing Artificial Neural Network (ANN) algorithms, based on models of biological neuron structures (see Appendix C). In speech recognition, the Multi-Layer Perceptron (MLP) is the architecture most commonly implemented.

Based on this model, Time Delay Neural Nets (TDNN) were first introduced for speech problems by Waibel et al. [2.1-8]. In such a model the basic unit of the neural network is modified, taking into account time delay constraints which are analogous to those used in Dynamic Time Warping.

The most widely used algorithms for pattern matching in ASR today are called Hidden Markov Models (HMMs). In these algorithms, a set of nodes is chosen for a set of phonetic or sub-word units. Three nodes, for example, could represent each phonetic unit [2.1-9]. The nodes are connected left-to-right with recursive loops (see Appendix D).

Recognition is based on a transition matrix of the probability of changing from one node to another and on a matrix, known as the output probability matrix, representing the probability that a particular set of features (e.g. MFCCs) will be observed at each node. These matrices are generated iteratively during a training process using speech from one or more speakers. These phonetic HMMs are then combined to form larger sets of nodes to represent words. Similarly, the sets of nodes representing words can be combined to form the legal sentences for the particular application.

During pattern matching each HMM model can be used to compute the probability of having generated the sequence of input spectra. This is done very effectively using the Viterbi algorithm [2.1-10] on the network of nodes used as the reference patterns. The result of the Viterbi algorithm is the total probability that the spectral sequence was generated by that series of HMMs using a specific node sequence. A different probability value results for every sequence of nodes.

For recognition, the above computation is performed for all possible phoneme models and all possible node sequences. Approximate search algorithms have been used to reduce the search computation without loss in performance. A commonly used technique known as beam search [2.1-11] is used to prune nodes that have low probabilities. The one sequence that results in the largest probability is declared to be the recognized sequence of phonemes/words/sentence.

It can be also shown that HMMs and ANNs can be linked together [2.1-12]. Such links have led researchers to integrate connectionist networks into a hidden Markov model speech recognition system. Then, it is shown that a connectionist network can be used as a probability estimator: in the classical HMM approach, topologies and probability density functions (pdf) are both chosen, initialized and estimated; in the approach described in [2.1-13], the topology of the HMM is still chosen but an MLP is dedicated to the output pdf estimator, through an iterative procedure, alternating between training the MLP and re-estimating the transition probabilities. The efficiency of this method has been shown through an evaluation on speaker-independent databases distributed by the Defence Advanced Research Projects Agency (DARPA). However, this technique remains dedicated to non-noisy speech recognition. Under adverse conditions, embedding preprocessing algorithms such as those described in paragraphs 2.1.4.1. and 2.1.4.2. should improve their performance.

2.1.2.4. Error Correction

It is likely that, for the foreseeable future, speech recognizers will always make some mistakes; after all, humans sometimes mis-hear what is said even under good conditions. In order to provide assurance that the voice input system takes the correct action in response to a spoken command, it is necessary for the user to monitor the recognizer output and have the means to correct any errors which have occurred. Feedback of the recognizer output may take several forms: visual, auditory, or implicit. Where a simple command (two or three words, without digits) is used to perform an obvious action such as changing display modes, no explicit feedback of the recognizer output is required; if the display changes as requested, the command was successfully recognized. If not, it is a simple matter to repeat the command. (There may be a problem regarding what actually did happen as a result of the mis-recognized command.)

More complex or critical commands will require the user to check the recognizer output before the command is executed. Feedback may be visual, (via the head-up display (HUD) or a special display), or auditory. Each has its advantages and disadvantages. Visual feedback is most reliable, but detracts from the eyes-out advantage of voice input. This is somewhat offset by the pilot being able to choose the time at which he looks at the feedback display. Auditory feedback leaves the pilot's eyes free for other tasks, but is transient and may be missed if the pilot's attention is distracted. It may also interfere with, or be overridden by, communications or auditory warning signals. A study on feedback modality [2.1-14] showed that providing both types of feedback gave the best performance on a voice input task and interfered least with a concurrent tracking task.

If an error is discovered in the recognizer output, means must be available to correct it before the command is executed. The simplest way to do this is to delete the whole command and repeat it; this will probably be the most effective way if the error rate is low. Alternatively, the vocabulary, syntax and system interface must provide a means to selectively delete and correct individual words in the command. Words such as "correction," "delete," or "insert" may be used to alter single words or digit strings, in which errors are most likely to occur. However, when the commands consist of only a few words, it is generally easier just to repeat the whole command.

After having decoded possible erroneous speech recognition, a dialogue can be used in order to correct the whole sentence, or a single word if the algorithm is accurate enough to localize the possible error inside the sentence. The problem is to design the interface between the man and the machine in such a way that the machine seems simple, or, at least, considerably less complex than it is. Moreover, the dialogue must be as generic as possible in order not to have to design

"ad hoc" dialogues from one application to another. So, the problem is to design a generic dialogue core which could be coupled to different applications. Figure 2.1-4 summarizes the previous explanations by describing the organization of such a dialogue system.

2.1.2.5. Acoustic Phonetic Decoding

Among all the methods developed during the last decades in speech recognition, one can distinguish "global" methods from "analytic" ones. Global methods recognize utterances by comparison to references, collected through an acoustical model of words. Dynamic Time Warping, Hidden Markov Models and Neural Networks are considered global methods.

Since spontaneous continuous speech production induces coarticulation effects, that have been studied by numerous authors, an analytic approach has been developed in order to localize and identify elementary entities during continuous speech production. According to the set of entities used, one can distinguish Acoustic Phonetic Decoding (APD) where elementary templates are the phonemes, or the diphones, or the syllables. Acoustic Features Identification (AFI) localizes and identifies phenomena that occur in speech production through acoustical characteristics such as voiced/unvoiced, plosive or not, fricatives or not, etc. Differences between ADP and AFI are small enough to consider them equivalent in this presentation. Even if the analytic approach is a potential method nowadays, global approaches still remain more efficient.

In order to control ASR, we must provide specific algorithms to detect speech recognition errors. One method consists of establishing acoustic phonetic decoding or speech feature extraction (see Figure 2.1-4) and analysis to be compared to the solution produced by the ASR. Such an approach is close to the techniques provided in analytic speech recognition, but the goal here is less ambitious than pure recognition: we only want to point out the main features of a sentence through a macro-phonetic classification (voiced/unvoiced speech, voiced/unvoiced fricatives, voiced/unvoiced plosives, ...).

Several accuracy levels can be taken into account: for example, if the pronounced utterance is "AUTO" and the ASR solution is "STOP", a voiced/unvoiced classification is sufficient to detect the error. But to separate "four" from "pour", voiced/unvoiced classification is irrelevant and a classification between fricatives and plosives is required. Such an approach could allow the detection of a large portion of speech recognition errors, especially in noisy applications where experiments show that ASR errors are, for the most part, irrelevant from an acoustic phonetic point of view. Such a strategy could not solve some difficult configurations without a perfect classification which would lead to a perfect ASR. But as long as ASR is not perfect, such an approach is relevant. Moreover, for military aircraft applications, such algorithms must be efficient in noisy environments.

As stated in section 2.1.2.2, wavelet analysis appears to be a relevant technical method to provide such algorithms. Wavelet decomposition is a powerful tool to analyze short-duration phenomena. After signal decomposition, entropy criteria-based algorithms provide relevant speech segmentation (see [2.1-15] and [2.1-16]). Moreover, in noisy environments, even in the case of correlated noise, the noise wavelet coefficients tend to be uncorrelated as the resolution and regularity levels increase. Rather than using entropy criteria-based algorithms, another method consists of applying new detection algorithms [2.1-17]. These algorithms allow fricatives and plosives detection (see [2.1-17] and [2.1-18]).

2.1.2.5.1. Speech Recognition Assessment

Speech recognizer performance is often expressed in terms of speech recognition rate. Speech recognition rate must be carefully used. In fact, the connected-word recognizer errors are generally assigned to three categories: *deletions*, where nothing in the solution provided by the recognizer matches with a particular word of the utterance; *insertions*, where a word recognized corresponds to nothing in the input; and *substitution*, where the word recognized is different from the corresponding word in the input utterance.

Each case is associated with a particular rate and performance is often obtained through a combination of these different rates and can be considered as a Word Recognition Rate

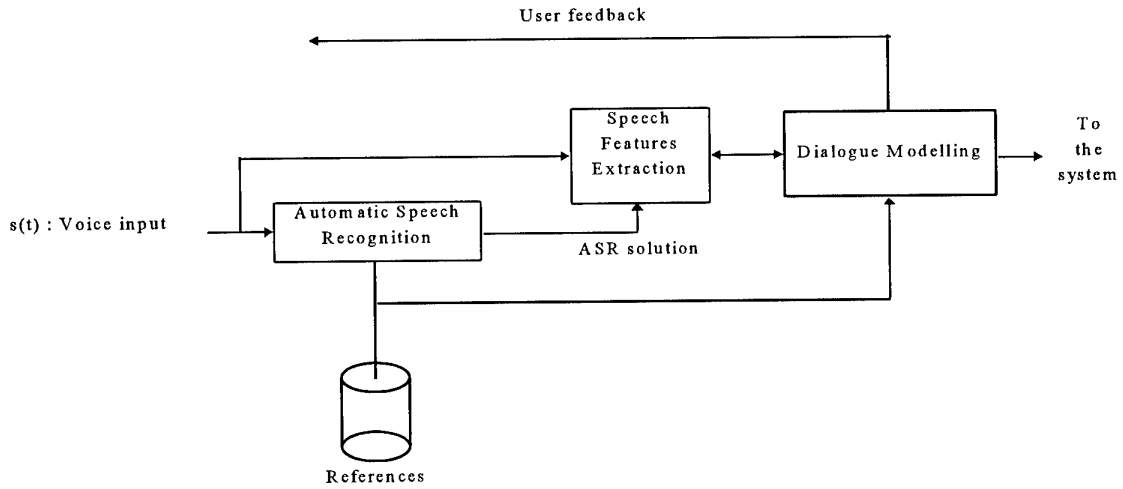


Figure 2.1-4 Diagram of Dialogue System

(WRR). On the other hand, it is possible to define a Sentence Recognition Rate (SRR) which is computed by considering that a whole sentence recognition is false as soon as there is only one word which has been misrecognized. It is clear that WRR and SRR are quite different. In a military aircraft cockpit, the commonly used Recognition Rate is the SRR. The SRR is more critical and based on the fact that, in aeronautical contexts, speech recognition errors imply consequences on the whole system. So, it appears very important to make speech recognition systems robust in order to avoid critical consequences on the system due a speech recognition error, as mentioned in the previous paragraphs.

2.1.3. APPLICATIONS TO DATE

2.1.3.1. Standalone

2.1.3.1.1. Fixed Wing

One of the first series of flight trials of speech recognition equipment took place between 1982 and 1985, on-board a BAC 111 civil airliner. This particular aircraft was a flying laboratory, based at the Bedford, UK, airfield of the Defence Research Agency. A speaker-dependent connected speech recognizer, the Marconi SR128, was used to control the displays, radios, and the experimental flight management system. Average recognition accuracy was over 95% on a vocabulary that was built up over a period to about 240 words. Some pilots found the system so useful that they used it as a normal part of their cockpit interface, even during trials of other equipment. The cockpit environment of such an aircraft is, of course, much less noisy and stressful than that of most military aircraft.

The U.S. Air Force, NASA, and the U.S. Navy conducted a joint program in the mid-1980's to flight test interactive voice systems in the fighter aircraft. The program consisted of laboratory and simulator testing prior to flight tests. Significant improvements in recognition accuracy were made during each of the three phases of the program. Speaker-dependent, isolated-word speech recognition systems were evaluated in the first two phases. A ten-word subset of that vocabulary was used in flight to control Multi-Function Displays (MFDs) in the cockpit of an experimental F-16 jet aircraft. The MFDs contained programmable switches, which selected pages of status information or control functions. The vocabulary words enabled the pilot to either address a particular page and then a particular function on that page, a specific function on a specific page, or select an aircraft master mode. These functions could be selected either manually or by voice. Performance was approximately 90% initially, but increased to the high 90's, for some pilots, by the end of flight tests. For those pilots with performance in the high 90's, speech was the preferred mode for interacting with the MFDs. Those pilots with performance in the low 90's preferred the manual mode of operation [2.1-19].

A Tornado GR1 has been used by the UK Defence Research Agency in two series of trials, in 1989 and 1993. The 1989 trials were aimed solely at collecting speech recordings in a cockpit environment representative of modern fast jets, but a recognizer was fitted to the aircraft to provide recognition feedback to the subject. These recordings were subsequently used to assess and optimize the Marconi ASR1000 flightworthy speech recognizer.

The second series of trials was intended to demonstrate the performance of the recognizer under realistic flight conditions. The navigator's main interface to the aircraft's main computer is via the Television Tabular display, known as TV-TABS for short. This has a small keyboard, but uses a complicated menu structure to access about 40 functions. Even quite simple operations may require many key presses, and the system is difficult to use and unpopular with the aircrew. A simple physical interface to the aircraft was possible, by breaking into the keyboard bus and making the recognizer output mimic key presses. This also allowed manual input to be mixed with voice input, even within the same command. Unfortunately, software reliability problems were encountered which could not be solved in time for the flight trials. Nevertheless, a total of 19 flights were made, with the navigator reading lists of command phrases and digit strings. An average recognition accuracy of over 95% was achieved. The final vocabulary size was 99 words and the syntax had a mean branching factor of about 15.

The U. S. Air Force has been conducting in-flight tests in recent years in a NASA OV-10 aircraft. These tests are to determine the present performance of speech recognition systems in the cockpit environment. The generic task selected was controlling communications and navigation functions. The vocabulary consists of 53 words or phrases. The system was tested in flight conditions of 1g and 3g and noise levels from 95 to 115 dB. Performance levels of better than 97% were obtained for 12 subjects in these conditions, using a commercially available speaker-dependent continuous speech recognition system.

The French Delegation Generale pour l'Armement (DGA) has been supporting studies and experiments dedicated to speech recognition since 1983. From 1983 till 1989, in-flight tests (mainly on Mirage IIIB but also on Rafale-A) have pointed out speech recognition systems limitations when used in a military aircraft cockpit. In light of these results, new algorithms have been developed and experiments in a centrifuge have been conducted in order to reduce the effects under adverse conditions (noise and G-load effects: see paragraph 2.4.1.1. and 2.4.1.2.). In 1989, a database was recorded during real flights under G-load on a Mirage IIIB aircraft. This database was used to evaluate speech processing and recognition algorithms performance (see [2.1-20] and [2.1-21]) before tests during real flights on the AlphaJet (described later in this section). The vocabulary was a restricted one, involving 36 words, allowing 9 linked words. The speech recognition algorithm was the preliminary version of TopVoice which is the Sextant Avionique Speech Recognition system (previously named DIVA). This speech recognition system is speaker-dependent, based on Dynamic Time Warping pattern recognition.

Two speakers took part in these experiments. Speaker 1 appears twice (Speaker 1a and Speaker 1b) because he used two different oxygen masks. The results are shown in Table 2.1-I.

Table 2.1-I Effects of G on speech recognition performance

Speaker - experimental conditions	Sentence Recognition Rates
Speaker 1a, 2g	100%
Speaker 1a, 4g	95%
Speaker 1b, 5g	96.4%
Speaker 1b, 2g	90%
Speaker 2, 2g	91.6%
Speaker 2, 4g	76.3%

Remark : As for the results which are described below (real flights in an AlphaJet), the recognition rate is a Sentence Recognition one, as described in section 2.1.2.6.: a whole sentence is considered as misrecognized as soon as there is only one recognition error, whatever the error is (deletion, substitution or insertion).

After these preliminary database experiments, TopVoice has been tested during flights on AlphaJet, the French training jet.

All flight configurations have been tested (speed from 200 to 450 knots, flight phases under G-load effects, low flight levels, real commands in context). Two syntaxes have been defined in order to take into account new functionalities involved in modern military fast jets (example: Rafale): the first one, to be used during cruise flight phases, involved more than 150 words and allowed sentences whose maximum

Table 2.1-II Sentence recognition rate, including all flights all speakers

First utterance	90%
First repetition (in case of error on the first utterance)	95%
Third utterance (in case of error on the first repetition)	97%

length was about 10 words; the second one was designed for flight under G-load, involved 25 real-time commands. These evaluations consisted of 80 flights, involving 15 different speakers and more than 10,000 vocal commands to recognize. The results are shown in Tables 2.1-II and 2.1-III.

These results show that the Sentence Recognition Rate (SRR) increases as soon as the pilot's attention increases.

These evaluations are broad enough to conclude about the main parameters whose influence is relevant :

- Noise is obviously one of these parameters, since the sentence recognition rate decreases as the noise level increases. Note that noise level increases not only during flight phases under G-load, but also as the speed increases. Despite this noise level, noisy speech processing avoids bad recognition rates and appears efficient.
- The microphone and audio circuitry must be optimized.
- The different parts of the syntax do not lead to the same results: systems commands, isolated words, digits are well-recognized, but international alphabet or numbers appear to be more difficult to recognize.
- Speech recognition remains very tied to speaker habituation as the results show. It depends on training phase quality and speaker vocal characteristics.

On the other hand, some effects are not so relevant as it seemed; specifically, G-load and Lombard effects.

The subjective conclusions of the users were that it appears easier to obtain data and parameters from the system when using vocal command. With a more and more complicated system to manage, vocal command is a relevant tool to decrease the workload, but vocal command must be controlled by a system able to detect recognition errors and to avoid disastrous consequences of speech recognition mistakes. Does it induce a dialogue between the pilot and the system, as soon as an error is detected? And first of all, how to detect erroneous recognitions? For these kinds of problems, refer to sections 2.1.2.4. and 2.1.2.5.

Such evaluations have shown the technical feasibility of speech recognition during flight and have identified some operational problems to solve, the main one being the system's ability to control its own recognition and to manage erroneous recognition.

The U. S. Air Force and the U. S. Navy are also conducting flight tests of speech recognition in the Joint Strike Fighter program. The application is a means of managing information and sensors. The vocabulary for this application is 12 words. The system has been tested in operational flight test conditions. Performance levels of 70% or greater have been obtained with three pilots. Two of the three pilots had performance of 90% or greater on several flights. The system tested is a militarized speaker-dependent isolated-word speech recognition system.

The European Fighter Aircraft EF2000 is a single-seat agile combat aircraft, planned to enter service about 2002. Speech input was included in the requirement from the beginning, and will be used for control of displays, radar, radios, target designation, navigation aids, and several other functions. Although test flying of the aircraft commenced in 1994, development of the speech recognizer module has not reached the stage of flight trials (at the time of writing). A commercially available speech recognizer has, however, been integrated into the cockpit simulator and used in the development of the man-machine interface. The reaction of pilots during the assessment program has been very positive. They regard speech recognition as essential to the safe and

Table 2.1-III Sentence Recognition Rate under G-load effects (5g)

First utterance	90%
First repetition (in case of error on the first utterance)	94%
Third utterance (in case of error on the first repetition)	98%

efficient operation of the aircraft.

2.1.3.1.2. Rotary Wing

The first in-flight use of ASR in a helicopter was in January 1981 [2.1-22]. These tests demonstrated that the most important problem to overcome for ASR in helicopter applications is the high noise level during flight.

The Day/Night All Weather (D/NAW) program in the UK, and the associated Covert Night and Day Operations in Rotorcraft (CONDOR) collaboration between the USA and the UK, are primarily concerned with advanced visual systems to allow rotary-wing operations to proceed in very poor visibility. The reliance on helmet-mounted displays can create a problem for the aircrew in operating switches and controls inside the aircraft, so voice input is an important adjunct to the visually coupled system.

Preliminary recognition trials in the DERA noise and vibration simulator have given good results. Mission-based trials in the Helicopter Mission Simulator in January 1997 compared missions flown with and without the use of voice input. Both pilot and commander had voice input, with different vocabularies. The pilot used about 25 words to control display modes and the radio altimeter (radalt); the commander's vocabulary of about 45 words controlled radios, map displays, transponder, and radalt. After the trial, the subjects, mainly operational Army aircrew with no previous experience of voice input, were strongly in favor of it, and considered it would offer a considerable enhancement to mission effectiveness. Following the simulator trials, a commercial speech recognizer was installed on the Lynx helicopter used for the D/NAW program at DERA, Boscombe Down, in the UK. Flight tests in late 1997 gave over 98% word accuracy.

Speech recognition has been tested on Gazelle, as a component of a Real-Time Digital Map Generator named MultiHélicare provided with graphical symbology overlaying capabilities. MultiHélicare is connected to the aircraft navigation system, to a voice command system and to a transmission system. The operator controls MultiHélicare with a joystick and the voice command system. The voice command system is TopVoice provided by Sextant Avionique, and which was described in the previous section.

Actions of the operator allow management of:

- the underlying map presentation,
- the overlaying symbology presentation,
- the loading and saving of the mission data,
- the aircraft navigation,
- the communications with another MultiHélicare system.

The main functionalities of MultiHélicare are:

- the friends/enemies tactical situation presentation and modification,
- the flight plans presentation and modification with automatic guidance,
- the dynamic terrain analysis by coloration and profiles display.

The syntax used for this application involved 67 French words and 2150 possible different sentences. The average length of the sentences is 3.5 words and the branching factor is 6.3. The SRR is over 95% during real flights, for any pilot. Moreover, tests conducted using an equivalent German syntax resulted in an SRR of over 98%.

The system has been used for several months during real flights, and it is very important to point out the subjective appreciation of the users who consider that the integration of speech recognition in a system such as MultiHélicare provides a tremendous amount of increased abilities, while decreasing the workload.

2.1.3.1.3. Space

Investigations into the utility of voice input/output (I/O) in the space shuttle were initially conducted in the mid 1980's [2.1-23]. The investigations centered around the control of the shuttle's Multifunction Cathode ray tube Display System (MCDS). This system is the main method the astronauts have for interacting with the five flight computers. Through the MCDS system, the astronauts do everything from reconfiguring the flight computers to checking the mission elapsed time. The MCDS has a 32-key oversized keyboard designed to be used with the bulky gloves of a space suit. A commercially available, speaker-dependent speech recognition system was used as an alternative to the keyboard. Similar applications of voice I/O are being considered for the space station as well [2.1-24].

An experimental voice command system was carried on shuttle mission STS-41 in October 1990, with the aim of collecting data on speech in microgravity conditions and to demonstrate the operational effectiveness of controlling spacecraft systems by voice. The recognizer was interfaced to the orbiter's closed-circuit TV system, which allows the astronauts to monitor the payload bay from inside the flightdeck. The speaker-dependent system used a vocabulary of 41 words to control the four TV cameras mounted in the payload bay. A very simple syntax allowed the cameras to be panned, tilted, focused, and allocated to one of two monitors. Two astronauts used the speaker-dependent system, with templates created on the ground before the mission. The system had the capability to retrain templates in space should the need arise. One of the astronauts experienced some initial difficulties due to the placement of his microphone, which was boom-mounted on a very light-weight headset. Once this was corrected, the system gave good results, and both astronauts were pleased [2.1-25].

There are plans for further assessment of voice input on future shuttle flights, possibly using it to control the manipulator arm. As a preliminary, the Canadian Space Agency included an experiment on simulated voice control of a robot arm during a short-duration space mission simulation. Four trainee astronauts spent seven days isolated in a hyperbaric chamber with workload and living conditions similar to those encountered in space, except for the gravity. The voice control tasks consisted of instructing a simulated 6 degree-of-freedom manipulator arm to grasp a ball while avoiding obstacles. The voice recognizer was simulated with the "Wizard of Oz" technique (see Glossary). The astronauts were not given a fixed vocabulary other than starting each command with the word "Viktor," but spoke spontaneously. Despite this, they used only 107 words in total between them, and only about 30 of these were common to all speakers. This experiment has helped to identify the vocabulary and

syntax most natural for the task and will contribute to further evaluation of voice input in space applications.

2.1.3.1.4. Command and Control

In the late 1980's, researchers at Boeing [2.1-26] investigated the utility of speech input/output (I/O) in the Airborne Warning and Control System (AWACS) man/machine interface. The present AWACS interface provides control and management of sensors through updating fields in tabular displays by inserting or changing alphanumeric values. This interface proves adequate for controlling one or two sensors, but it begins to overload the operator as more sensors are added. Operator tasks were analyzed to identify those thought to be best performed by speech I/O. Based on these functions a vocabulary and grammar were developed for a commercially available speech recognition system. This system demonstrated the effectiveness of voice I/O for several functions including fuel updating, committing fighters, and tactical broadcast control. The studies identified several features required of speech I/O in the AWACS operational environment. In the AWACS environment, the operator is under stress, there are multiple voice communications occurring in the background, and few I/O errors (speech input not recognized, speech output not heard by the operator) can be tolerated.

Another application of speech recognition in the late 1980's was training of air traffic control (ATC) trainees in the use of the correct ATC technology and phraseology [2.1-27]. The concept is that the trainee/speaker runs through a set of ATC scenarios. He speaks sentences intended to be appropriate to the scenario. The system provides feedback, identifying items and places where the vocal behavior of the trainee must be altered. The trainer used a commercially available, speaker-dependent continuous speech recognition system.

2.1.3.2. Integration

One knowledge source that has largely been ignored by ASR researchers is the face of the speaker. Lip-reading clearly meets at least two criteria which have been used to select techniques to improve ASR systems: it mimics human perceptual processes, and it yields information which is not always present in the acoustic signal. The use of visual information in the form of a video sequence of a speaker's face provides a useful knowledge source which improves the performance of ASR systems, especially under poor acoustic conditions or when extraneous signals (such as other speakers) are present. Lip-reading promises to broaden the range of applicability of speech recognition. The combination of ASR and lip-reading is also referred to as audio-visual speech recognition or speech-reading.

It is well-known that hearing-impaired humans use lip-reading as an important, even primary, source of information for speech perception. It is also well-known that persons with normal hearing use lip-reading as a redundant information source when listening conditions are difficult, such as crowded, noisy rooms or windy days. Situations such as trying to understand a foreign language, or one's own native language spoken by a foreigner, even in a quiet environment, are also difficult. Lip-reading can help a person to understand speech in such conditions [2.1-28 - 2.1-31].

Clearly, the use of lip-reading increases the robustness of the human speech perception process. In acoustically noisy environments, it offers a channel which is unaffected by processes which contribute to the acoustic noise. Lip-reading provides an independent dimension upon which to base linguistic decisions, and many phonemes which may be said to be "close" acoustically are "distant" visually [2.1-29].

2.1.3.2.1. Overview of Approaches

The goals of video processing are to preprocess a sequence of images and extract features suitable for recognition. The form of the images can vary substantially, and the features to be extracted depend on the application. Images could be in full color, grayscale, or black and white, showing the full face or just the mouth region, frontal, profile or from an arbitrary and possibly changing orientation.

For a lip-reading system to be employed in a real-world application, the algorithm must be robust to variations in illumination, skin color, size and distance of the talker, head rotation, facial hair, makeup, etc., and thus overall lightness and contrast compensation may be needed as a first step in video processing.

The type of feature vectors used differs from system to system. Low-level features, such as raw pixel values, the area of the mouth cavity, separation of the lips, and so forth, can be processed to give higher-level and more abstract features. Some systems use physical measures such as width, height, jaw rotation as features, whereas other systems may use the entire image possibly after some simple pixel-level filtering operations.

Feature location. A few systems are used where the talker's head is fixed relative to the camera, as, for example, military helmets. However, if the talker is allowed to move around freely, the first task is to locate the talker's head and regions such as the mouth and jaw. Various techniques are used to search for the head. One technique is based on color or, more specifically, hue or chrominance, to find a face in the image. Skin hue is surprisingly constant across talkers, and even across races, despite the fact that the lightness can vary significantly. One may need to add additional constraints so as to reject hands or other skin-colored objects.

One can also use edges and intrinsic values to find possible candidates for eyes and mouth. Facial features are usually characterized by high edge content and low intrinsic values.[1] Since the spatial relationship between eyes and mouth is fixed, erroneous candidates can be eliminated. This approach suffers from occlusions of the mouth by facial hair, however, and glasses can make detecting the eyes difficult.

If the camera can be constrained to be positioned below the face, the nostrils become a very distinct feature because they are effectively black and not occluded by any facial hair or by glasses. Unfortunately, one does not always have control over the camera position. In desktop video the camera is typically mounted on top of the monitor primarily because it leads to a more familiar and natural image of the talker.

Motion can also be used to find a talker. In many situations, the talker is the only moving object and within the face the mouth will move the most. Comparing subsequent frames of

[1] The intrinsic value of a pixel is the ratio of its intensity to the average intensity in a large neighborhood. Thus, it is an estimate of the "lightness" of an object independent of the illumination.

an image sequence, it is possible to detect areas where motion has occurred. This approach will require supplementary information, such as expected size or general location, if there are other persons or objects moving in the background.

A successful lip-reading system may need to combine several or all of the above techniques to obtain a reliable estimate of the head location, depending upon the application environment. In the most general application, where the face can be of arbitrary size and orientation, complicated model-based face-finding techniques may be required.

Feature extraction. Conceptually the simplest feature extraction method is to present the entire raw image to the recognizer, effectively making every pixel value a feature. While this approach guarantees that no information is lost, it does not reduce the data in any significant way. Because the required number of training patterns increases with the number of free parameters in the recognizer (and hence with the number of features), such pixel-based methods may require an unacceptably large number of training patterns, and perhaps a very long training time.

A deeper criticism of this method is that it is not robust to variations in illumination; if the amount, color or direction of the illumination changes, all the pixel values change. Likewise, if the camera is not centered on the same position in the face, the pixel values change, severely degrading recognition.

At the other extreme, one might seek a complete 3-D model of the mouth and jaw region that takes into account the various muscles at play, elasticity of the skin, and so on. Usually only a small set of parameters is required to completely describe a model; thus significant data reduction is achieved. The parameters of the model are then likely, but not guaranteed, to be linguistically relevant.

Examples of lip-reading systems. Finn and Montgomery [2.1-32] performed studies on optical-only recognition of consonants pronounced in an /a/-C-/a/ context. Visual processing was simplified by the use of 12 highly reflective dots placed around the mouth at specific locations. Fourteen distance measurements were derived from the dot positions, and these values were sampled every 30^{th} of a second. Training was accomplished by storing a template for each consonant, and recognition consisted of determining the template which minimized a weighted Euclidean distance metric in the 14-dimensional parameter space. Utterances were truncated to a common length, so no time warping was required. Again, this method was tested on only a single speaker. Each nonsense word was spoken twice: once as a template and once to test the system. *A posteriori* optimization of the weights gave 87% recognition accuracy of the phonemes. No attempt was made to merge the system with an audio recognizer.

An obvious disadvantage of this method is the requirement that the speaker wear reflective dots on his or her face. No time warping was used; only a simple time shift was used to align sequences. For this particular experiment, this may have been appropriate. The system performed surprisingly well in discriminating voicing. The authors conjectured that this was because it was sensitive to subtle timing differences (voiced consonants had slightly longer duration). Sequences were aligned at the point of maximum closure of the mouth; thus, the consonants to be discriminated were guaranteed to be aligned with each other. This alignment strategy cannot be extended to natural speech recognition, so it is very doubtful that their results for consonant discrimination can be generalized to natural speech, either.

Mase and Pentland [2.1-34 and 2.1-35] used optical flow, computed in four windows near the mouth, to extract parameters for visual speech recognition. The parameters are weighted sums of certain motion components in the four windows, and correspond to changes in the width and height of the mouth. They argue that this approach should result in a speaker-independent system, for the reason that the flow fields will reflect the action of the major muscle groups involved in speaking. While this may be true, there is also a great deal of important information which is discarded, such as the positions of visible articulators internal to the mouth. Linear time warping was used to match test utterances to previously stored templates. No attempt at integrating audio information was made, except as noted below.

Speaker-independent experiments were carried out with a ten-digit, continuous-digit task and accuracy varied from 50% to 100%, depending on the speaker. Results improved to 75% to 100% when the audio signal was used to determine the time of onset of speech. Interestingly, of the two speakers who scored 50%, one was bearded and one was a native Japanese speaker. The relatively low score of the Japanese speaker is to be expected of any English language ASR system, but the poor performance of the bearded speaker may indicate a more serious problem. The sample size (only eight test utterances of three to four digits each) was too small to draw any firm conclusions.

Goldschen [2.1-35] developed a completely visual speech recognition system. The system uses 13 dynamic features to characterize the oral cavity, unlike the static features used by other researchers. Phonemes that appear optically similar were merged into visemes. The visemes were objectively analyzed and discriminated using HMMs and clustering algorithms. These visemes are consistent with the phoneme-to-viseme mapping discussed by most lip-reading experts. HMMs were trained to recognize, without a grammar, a set of 150 sentences from the TIMIT database having a perplexity of 150, using visemes, trisemes (triplets of visemes), and generalized trisemes (clustered trisemes). The systems achieved recognition rates of 2%, 12.7%, and 25.3%, respectively. This system is the first to achieve continuous speech recognition with visual input.

2.1.3.2.2. Examples of speech-reading systems

The literature on automatic speech-reading was quite limited until recently. The first serious work in the area was performed by Petajan [2.1-36 - 2.1-38] in 1984. Others to research the area include Yuhas et al. [2.1-39 - 2.1-41], Stork et al. [2.1-42], and Silsbee [2.1-43]. Research has accelerated in the last several years with conferences and workshops now held regularly on this subject [2.1-44 - 2.1-46].

Petajan built a system consisting of two parts: a commercially available, audio-based recognizer and a computer vision system. The computer vision system used a simple thresholding technique to extract information about the shape of the acoustic opening. Thresholding of the mouth image yielded a blob representing the open mouth. The height, width, circumference, and area of the opening were recorded. A linear time warping algorithm was used to match samples

to exemplar templates. The resulting score was used to choose between the top two candidates picked by the audio system.

Petajan's system was tested using a single speaker (the author), with three tasks: the 10 digits, the 26 letters, and a 100-word vocabulary. Lip-reading improved recognition in all cases, compared to the audio recognizer alone: from 95% to 100% for the 10-digit task, from 64% to 66% for the letters task, and from 65% to 78% for the 100-word task.

One major problem with this system is the thresholding technique used; it seems unlikely that the same threshold would work well for all faces; indeed, dark skin or a beard would probably interfere with the parameter extraction. It is also sensitive to lighting conditions; the author was very specific about how lighting was to be set up. The linear time warping algorithm allowed only simple dilation and contraction of time, which cannot encompass all of the natural variation of speech.

In addition, although integration of the two streams was successful, it was clearly not optimal. The percentage of the time that the correct word was ranked first or second by the audio recognizer places a hard upper limit on the performance of the combined system. Also, the relative confidence of the two subsystems made no difference in the final outcome. Even if the top two audio scores differed widely, indicating a high degree of confidence, and the corresponding visual scores were nearly identical, the visual subsystem would be allowed to make the decision between those two words. In that case, performance of the combined system would certainly be worse than that of the audio system.

Yuhas et al. used a neural network (NN) to integrate audio and visual information. They presented 20 x 25 pixel images of the mouth area directly to a feed-forward NN which was trained with a modified back-propagation algorithm. The data set consisted of single images of a speaker uttering nine different vowels, with several samples of each.

Sensory integration in the Yuhas system occurred at an earlier stage than in the Petajan system. The NN was not trained to directly classify the test utterances, but rather to produce an estimate of the power spectrum of the audio signal, based on the visual signal. This was additively combined with the power spectrum of the audio signal and the result presented to the classification algorithm.

Tests were performed using a single speaker over a wide range of signal-to-noise ratios (SNR). Again, *a posteriori* optimization of the relative weights of the two spectrum estimates was used. At high SNR (24 dB), the addition of the visual estimates made essentially no difference, while at very low SNR (12 dB), it improved accuracy from 11% (chance performance) to over 50%. This system was only capable of dealing with single frames, and thus was not adaptable to real-time implementation.

Stork et al. investigated several neural network architectures for the integration of audio and visual information. Visual input consisted of the positions of 10 reflective dots which had been affixed to the speaker's face. They chose a 10-letter task for their experiments, using five talkers. Their results were not presented in numerical form, but they indicated a performance improvement due to the visual information.

Silsbee developed a system that uses lip-reading to augment an ASR system. The system uses HMMs and Vector Quantization (VQ) on both the audio and visual streams. The audio subsystem uses PLP to parameterize the speech signal. This is then VQed and input to a speaker-dependent HMM. The lip-reading subsystem uses a modified-VQ algorithm that is sensitive to important visual speech perception features such as lip separation and visibility of the teeth. These features are also input to a speaker dependent HMM. The two subsystems word scores are linearly combined to make the classification. The system has been tested using three tasks: discrimination of 22 consonants in an identical context, discrimination of 14 vowels in an identical context, and discrimination of a 500-word vocabulary. The system requires no markings on the face, and should be reasonably resistant to minor lighting variations (this has not been thoroughly tested). Since the system uses HMMs, it can easily be extended to natural (continuous) speech. Performance improvements were shown on the consonant and vowel tasks.

In more recent work, Silsbee and Su [2.1-47] investigated different methods to integrate the audio and visual HMMs. Results showed that an off-line adaptive weighting method was always as good as or better than that obtained by any of the fixed weightings they tried. Current work is investigating an on-line method to adapt the relative influence of the audio and visual information.

Hennecke, Stork, and Prasad [2.1-48] have developed a system that uses an automatic procedure to locate the face and mouth in a color image. They use a model-based procedure incorporating deformable templates to match the current image of the lips. The resulting parameters of the template are used as the features for the lip-reading sub-system. The features from the speech and lip-reading sub-systems are concatenated and used as the input to an HMM for recognition.

In a four syllable recognition experiment, the syllables /da/, /fa/, /la/, and /ma/ were recorded from 10 talkers, and repeated five times at an SNR of 10 dB. Of the 200 utterances, 160 were used for training and 40 were used for testing. Using audio and video alone, error rates of 57% and 62% respectively were obtained. When combined, the error rate dropped to 41%.

Bregler et al. [2.1-49] have developed a hybrid speech-reading system that uses manifold learning, neural networks, and HMMs. The manifold learning is used to track and extract the lips and for providing the feature for recognition. The system uses a neural network/HMM hybrid with features based on Relative Spectra (RASTA-PLP) for the speech recognition. An HMM is used for the visual classifier.

The system was tested using a database of six speakers spelling names or saying random letter sequences. Each utterance was 3-8 letters. The acoustic signal was distorted with background noise or crosstalk from another speaker. Results compared the acoustic-only recognizer with two versions of an audio-visual recognition system. The first visual recognizer used 10 coordinates from the lips determined by the manifold learning procedure, and the second used an additional 10 delta features of the lip coordinates. For clean speech no significant improvement was obtained. For noise-degraded cases of 10 dB and 20 dB

SNR significant improvements were obtained. The largest improvements were obtained in the crosstalk experiment (15 dB SNR). The system which used the delta visual features consistently outperformed the other systems in the degraded conditions.

All of these studies reported some degree of success. In particular, all showed improved results when audio speech recognizers were supplemented with the lip-reading systems, especially in acoustically degraded conditions. These studies all appear to show that there is useful information in the sequence of images of a speaker's face. All, however, are very small-scale studies in terms of vocabulary size and size of the data set. Also, some systems required "artificial" assistance in some form (such as reflective dots on the speaker's face).

2.1.3.3. Monitoring

It is common experience that many aspects of a person's physical or emotional state may be detected from the sound of his voice, but detailed knowledge relating changes in measurable parameters of speech to particular kinds of stress is very limited. Two major problems are that stress can be very difficult to define, and that individual reactions to it may vary over a very wide range. However, given that humans can classify others' emotional states from their voices with some degree of accuracy, it must be possible, at least in principle, to automate the process.

Physical stresses, such as G-force and vibration, have relatively well-defined effects on speech production, because they act directly on the vocal apparatus without the mental interpretation that intervenes in the case of many other stressful stimuli. Nevertheless, the effects are still dependent on the subject's level of training and experience under the particular stressor. In practice, the physical conditions in an aircraft can be measured accurately and reliably by physical sensors, so there is little need to use voice monitoring in this way. It may, however, find an application in accident investigations when there are no physical measures available.

There is considerable psychological literature on the effects of stress and emotion on the voice [2.1-50], but most of the practical interest has been associated with space flight. Given the isolation, danger and expense of space missions, monitoring of the astronauts' state may be crucial to avoiding a disaster. Stress levels can be determined by means of physiological measures, but the associated sensors and wiring will be inconvenient in the confined cabin of a spacecraft. Also, where the astronaut is required to work outside the spacecraft, they will complicate the process of donning the pressure suit, and will require extra telemetry bandwidth. There has therefore been considerable interest in using the voice to monitor the state of the astronaut.

Many experiments have found changes occurring to voice parameters under stress conditions, but there have always been very large differences in responses between speakers. Some of this variance is associated with different reactions obtained from different personality types, and also with gender. It is possible, though, that more consistent reactions would be obtained from a group as highly selected and trained as astronauts are. Even so, a reliable "voice stress monitor" seems a long way off.

2.1.4. APPLICATION PROBLEMS

2.1.4.1. Noise

One problem that all current systems share is that their performance degrades significantly as conditions depart from the ideal noise-free case. Recognition errors can increase dramatically in the presence of noise, which can come in a variety of forms. All real-world applications are subject to interference from noise, whether it is due to fans in an office environment, vehicle engines, machinery, or even other voices in the background. The usefulness of an ASR system is limited by how well it can handle such problems.

Recently, there has been some success in this area. Two areas of focus have emerged: improved modeling of the acoustic phenomena and the use of additional knowledge sources such as improved language models or prosody.

Modeling of acoustic phenomena has focused primarily on reducing the effects of noise on speech. To combat this kind of effect, various speech enhancement techniques have been investigated. These have resulted in error reduction in ASR systems of around 35% to nearly 100%, depending on the task and the amount of noise [2.1-51 - 2.1-57].

Another area of focus has been that of representing the acoustic signal in ways that relate to the human auditory system, since humans perform very well at speech recognition [2.1-3, 2.1-58 - 2.1-60]. Since a very large reduction in data dimension and data rate takes place between the sampling of an acoustic signal and the representation of that signal which is used in the recognition algorithm, it is critical that the reduction take place in a manner which preserves the important information. Filters and signal processing methods are designed which mimic processes that occur in the inner ear and the brain. It is thought that the information extracted by these techniques is likely to be linguistically relevant and more robust to effects of noise.

Among these methods, some have been implemented by speech/noise discrimination [2.1-20 and 2.1-21]. Noise robustness associated with speech/noise discrimination has been involved in a flyable speech recognizer which has been tested during flights on helicopters and fast jets (see sections 2.1.3.1.1. and 2.1.3.1.2.). As described in [2.1-20], preliminary experiments on a database recorded during flights of a Mirage IIIB (see section 2.1.3.1.1.) have shown that speech detection alone was able to improve the speech recognition rate, even under G-load effects. If noise cancellation is added, there is an additional speech recognition rate gain, but which is lower than the gain due to detection.

It is quite obvious that speech/noise discrimination improves the speech recognition rate. It is more difficult to understand why speech detection is so important. In fact, a pilot uses a push-to-talk (PTT) in order to give a voice command to the system. The pilot's PTT is not perfect and, in most cases, is longer than the real speech duration. Speech recognition algorithms begin the recognition process during a noisy pause; this can induce bad choices in the syntactic tree structure. Owing to accurate speech detection through speech/noise discrimination, such a phenomenon can be avoided.

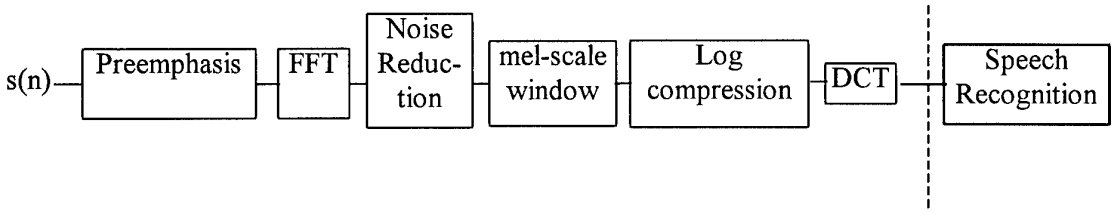


Figure 2.1-5 Speech Processing

Speech/noise discrimination and noise cancellation are closely related problems, because noise cancellation algorithms need statistical and spectral information about the background noise of interest. The noise can be considered stationary during a vocal command, but from one vocal command to another, its characteristics (for example, its level) can change. So, noise cancellation requires the detection of noise in order to adaptively extract its spectral and statistical parameters. The ability to discriminate speech from noise enables the calibration of noise cancellation algorithms. The result of such an approach, is described by Figure 2.1-5 which depicts the whole processing chain. Noise cancellation is assumed to be performed by Wiener Filtering.

This principle has been tested on a data base recorded during real flights under G-load on Mirage III B (see section 2.1.3.1.1.). The results obtained are described in Table 2.1-IV, where the nomenclature is the following one:

- PTT: results obtained when the pilot's original Push-To-Talk is used in order to define the beginning and the end of the utterance
- SD: results provided with Speech Detection alone
- SD+NC: results provided by the complete algorithm (Speech Detection and Noise Cancellation)
- PWB+NC: results obtained with a Perfect Word Boundary Detection and Noise Cancellation

In each column of Table 2.1-IV, the number of errors and the number of utterances are given, as well as the recognition rate: for example, 12/30 (60%) indicates 12 errors in 30 utterances, and the recognition rate is then 60%.

Figure 2.1-6 illustrates the Noise Cancellation efficiency of such an approach on the utterance "Donne Page Hydraulique."

Woods asserts that "there is not enough information in the acoustic signal alone to determine the phonetic content of the message". Humans rely on other knowledge sources to help constrain the set of possible interpretations. These are useful for machine recognition as well, and indispensable for many tasks. The most important knowledge source is grammar. Grammar places strong restrictions on the set of words which can follow or precede a given word [2.1-61 - 2.1-65]. Others include prosody (information contained in the rhythms and pitch variations of speech) [2.1-66, 2.1-67] and focus (constraining the vocabulary to the topic of a "conversation") [2.1-62]. The use of visual information has been covered in section 2.1.3.2.

2.1.4.2. Stress

Stress is a rather ill-defined concept, covering a multitude of generally threatening conditions. Many of these have elements in common, particularly those that activate the autonomic nervous system, but the external stimulus is always subject to a greater or lesser degree of mental interpretation which results in individual reactions varying widely. In addition, training and experience can have a large effect on how well individuals cope with many kinds of stressors. The effects of stress are usually apparent in the voice, and hence affect the performance of speech recognizers. The problem, as always, is that the conditions of use are different from those under which the recognizer's models are trained. This mismatch is largely unavoidable as it is usually impractical, expensive or unethical to subject a user to such stresses in order to train the recognizer.

Table 2.1-IV Speech recognition rates with/without speech/noise discrimination and with/without noise cancellation

Environmental conditions	PTT	SD	SD+NC	PWB+NC
speaker 1 - 2g	5/36 (86.1%)	2/36 (94.4%)	0/36 (100%)	0/36 (100%)
speaker 1 - 4g	4/60 (93.3%)	5/60 (91.6%)	3/60 (95%)	3/60 (95%)
speaker 1 - 5g	3/28 (89.2%)	4/28 (95.1%)	1/28 (96.4%)	1/28 (96.4%)
speaker 1 - 2g	12/30 (60%)	6/30 (80%)	3/30 (90%)	2/30 (93.3%)
speaker 2 - 2g	39/48 (18.75%)	11/48 (77%)	4/48 (91.6%)	2/48 (95.8%)
speaker 2 - 4g	53/55 (4%)	23/55 (58%)	13/55 (76.3%)	8/55 (85.4%)

2.1.4.2.1. Physical stress

Physical stresses may be classified under four main areas: the force environment, auditory distraction, the thermal environment, and personal equipment. For aircrew, the major factors in the force environment are G-force, vibration and pressure (cabin pressure or pressure breathing for G protection). Some experiments have shown that highly trained and experienced personnel can speak relatively normally at up to 5g with only about 5% loss in recognizer performance, but inexperienced subjects may suffer 30% loss in performance at lower G-levels. Vibration is the predominant problem in rotary wing aircraft. Dominant frequencies from the main rotor lie in the range of 5-30 Hz; typical resonant frequencies of body structures of the torso and head also lie in this range. Pressure breathing for G-protection involves increasing the pressure of the breathing gas by as much as 50 mmHg or more. This inflates the vocal tract and makes speaking difficult.

Some studies have been conducted in order to determine not only the speech recognition rate degradation due to G-load effects, but also in order to point out efficient speech processing able to balance these degradations owing to an analysis of speech production alterations under G-load [2.1-68].

These studies are based on experiments in a centrifuge, involving six pilots whose mean age was 30. Through different signal analysis tools (pitch detection, short-time Fourier transform, Multiresolution analysis, Principal Component Analysis), it has been possible to study speech production modifications at different G-load levels (1.4g, 3g, 6g). Even if these tools point out some typical phenomena correlated with identified physiological mechanisms, it remains difficult to integrate such considerations in a speech recognition system since these phenomena remain variable and hazardous.

However, this study points out that detecting speech from

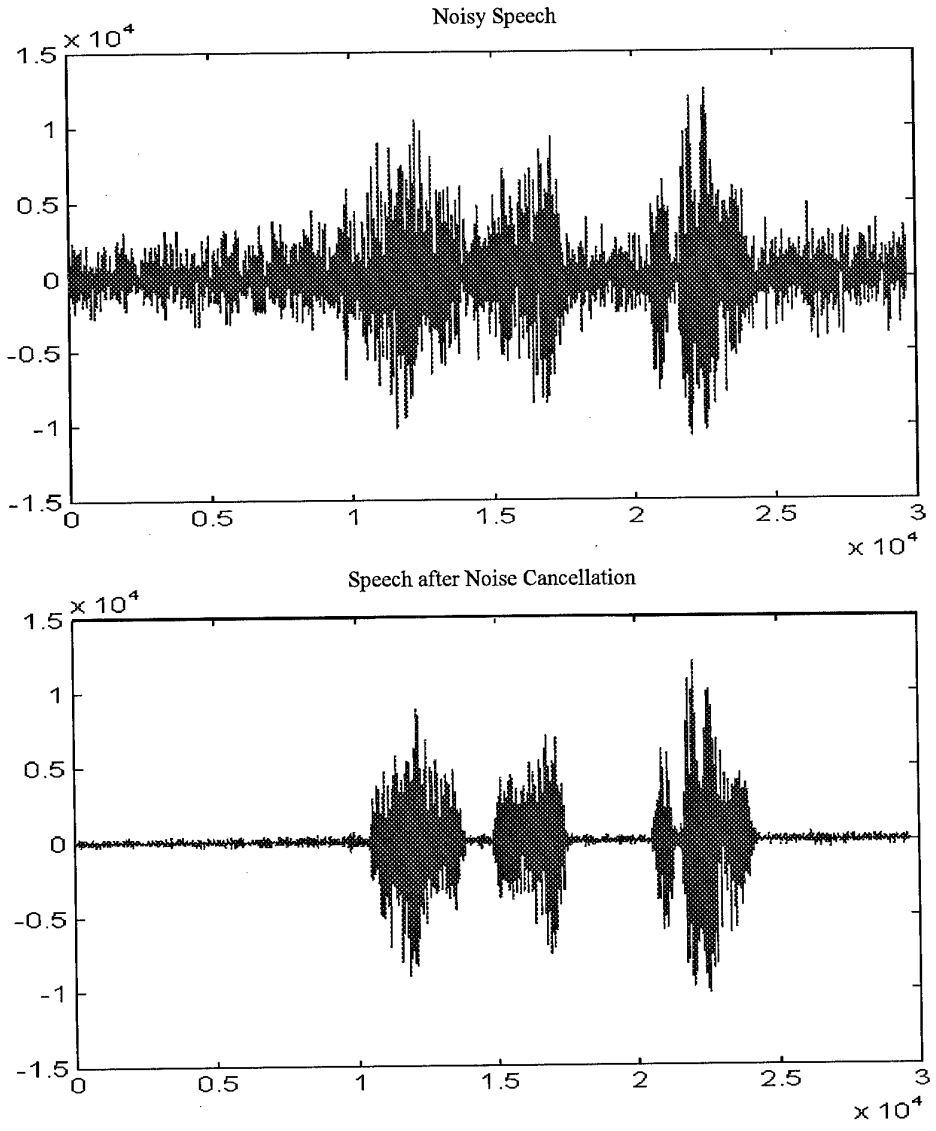


Figure 2.1-6 Noisy speech (top) and after Noise Cancellation (bottom)

pause and reducing the vocabulary complexity were two relevant means in order to get acceptable speech recognition, even under G-load effects. Speech detection principles and influences on the speech recognition task are, under G-load effects, the same as those described in section 2.1.4.1. Reducing syntax complexity is not simply a trick but fits with section 2.1.5. recommendations and human physiological abilities, since it becomes really difficult to speak clearly and naturally, except for highly trained personnel. Finally, from an operational point of view, the number of required speech commands decreases quickly.

In order to take into account each environmental parameter whose influence is relevant, some studies have been conducted in order to determine the speech production modifications due to combined stress (workload, noise, G-load, positive pressure breathing). A database has been recorded in a centrifuge and the data are currently being processed. Such an analysis should provide some constraints that future speech recognizers will have to respect in order to be relevant under complex fast jets environmental conditions. Such a theme is close to the current NATO working group IST/TG001 (formerly RSG10) dedicated to state-of-the-art speech processing.

Noise levels are high in modern military aircraft, often 110-115 dB SPL. Hearing protection is improving, but many aircrew can still expect to be subject to levels of around 85 dBA for the duration of the mission. Short-term effects can be compensated for by training the recognizer under similar noise conditions, but these noise levels can also create mental fatigue over a period. Other auditory stressors include auditory warnings and voice communications which add to the total noise dose and may carry distracting or anxiety-causing information.

The thermal (i.e. temperature and humidity), environment of military aircraft is in general not too extreme, but may become so in the event of a failure or battle damage. At present, there is not much detailed knowledge about the effects of temperature on the voice.

Personal equipment includes clothing, helmet, oxygen mask, NBC protection, and safety harnesses. These may restrict movement in various ways or apply pressure to the body. The oxygen mask is a special case, in that it is intimately involved in speech production. The effect that the mask has on the speech spectrum is considerable, but is not a stressor as such. The mask may also constrict jaw movement, add to fatigue, and, over a long period, apply painful pressure to the face.

2.1.4.2.2. Emotional stress

Emotional stresses may be classified under the general headings of task load, mental fatigue, mission anxieties and background anxieties. Task load arises out of the immediate demands of the mission on a crew member, requiring him to absorb information, make decisions and take actions. Mental fatigue affects general alertness, and may arise from loss of sleep, physical fatigue or boredom. Mission anxiety arises out of threatening situations that occur in the course of the mission. As well as the obvious threats arising from enemy action, this also covers social aspects such as the weight of responsibility and difficulties in interactions between crew members. Finally, background anxieties covers aspects of domestic, career and health worries which do not arise out of the mission itself but can have a significant impact on aircrew performance.

2.1.4.3. Accent

It is well known that speaker accent is one factor that degrades the performance of present-day speech recognition systems [2.1-69]. This is a problem that occurs no matter the target language on which the recognizer was trained [2.1-70]. Several approaches to this problem are to first identify the accent [2.1-71 - 2.1-73] and then use a recognizer trained on that accent [2.1-74, 2.1-75], select an appropriate language model [2.1-76], or adapt to the accent/speaker [2.1-77]. Each of these approaches has trade-offs in terms of training complexity.

Degradation in recognition performance due to accent is a concern in commercial applications running on the telephone network and on personal computers. It is also a concern in military applications with the now-common multinational forces and in air traffic control. This area will get increased attention because of the significant benefits that will be derived in commercial applications. The unique military aspects will be the effects on speech recognition performance with combinations such as accented speech in a stressful, high-noise environment.

2.1.5. REQUIRED ENHANCEMENTS AND PROGNOSIS

Speech recognition performance for small and large vocabulary systems is adequate for some applications in benign environments. Any change in the environment between the training and testing causes a degradation in performance. Continued research is required to improve robustness to new speakers, new dialects, and channel or microphone characteristics. Systems that have some ability to adapt to such changes have been developed [2.1-78, 2.1-79]. Algorithms that enable ASR systems to be more robust in noisy changing environments such as airports or automobiles have been developed [2.1-80 - 2.1-83], but performance is still lacking. Speech recognition performance for very large vocabularies and large perplexities is not adequate for applications in any environment. Continued research to improve out-of-vocabulary word rejection in addition to the above-mentioned areas will enable larger vocabulary ASR systems to be viable for applications in the future.

An answer to the problem of the user having to remember a large vocabulary is to make the system capable of understanding any command, however it is phrased. The user can then speak naturally, using whatever form of words comes to mind at that instant. This removes the workload associated with having to remember which words are valid. Such systems are often called "speech understanding" systems.

The simplest systems use word spotting techniques. For example, to select a radio frequency with a finite state syntax, the pilot may have to say, "RADIO VHF HEATHROW APPROACH." A natural language system could accept "GIVE ME HEATHROW APPROACH ON VHF" or "SELECT VHF, ER, I WANT HEATHROW APPROACH." The system needs only recognise the words "VHF," "HEATHROW," and "APPROACH" to infer that the VHF radio should be tuned to that channel. Words which are not a good match to key words in the vocabulary are matched to a

so-called "garbage model," which approximates the long-term speech spectrum. Another approach is to attempt to recognize all words spoken, then pick out the key words from the resulting word stream. The overall error rate may be relatively poor, but providing that the key words are recognized correctly, useful output may be obtained.

Many speech understanding systems attempt to make use of several different areas of knowledge about the speech and the situation in which it is being used. Starting with a parametric representation of the speech signal, hypotheses are formed about possible phone sequences. Phonetic and phonological knowledge is used to provide constraints at this level. From these sequences, higher-level hypotheses are formed about possible word sequences using syntactic, prosodic and lexical knowledge. Constraints may be added from knowledge of the application and the current situation, until finally a single sentence emerges. A reliable natural language interface may be some way off, but is a prime goal for research in speech recognition.

Integration of other alternative controls with speech-based control have recently been attempted and will require further technology development. The first and foremost requirement to advance the field of lip/speech-reading is the collection of a high-quality labelled database. This is needed both for training and for comparing the accuracy of different methods. The next requirement is exploration of sensory integration and recognition methods that are more appropriate for the unique requirements of speech-reading such as the different time scales between the audio and visual channels, need for rate invariance, etc. Additionally the system needs to adapt in real-time to changes in overall noise level.

Use of speech-based control as a supplement to conventional controls is becoming common. For example, a system designer can make cockpit radio frequency selection or multi-function display operation accessible with speech-based as well as conventional control systems. The user could choose to use the speech-based system when appropriate. An analogy is the availability of both keyboard and mouse functions for cursor positioning in a modern personal computer system. Users will choose one or the other depending on the nature of the task, hand location and personal preference. One key issue that must be addressed is the ability to operate speech-based controls in multi-task environments. Some research has investigated the effect of task loading and other physical stressors on speech and its resultant impact on speech recognition performance [2.1-84, 2.1-85]. Continued research is needed to reduce the impact of these factors.

2.1.6. REFERENCES

2.1-1 Flanagan, F. L., "Speech Analysis, Synthesis and Perception", Springer Verlag, New York, 1972.

2.1-2 Deller, J. R., Proakis, J. G., and Hansen, J. H. L., "Discrete-Time Processing of Speech Signals", Englewood Cliffs, USA, Macmillan Publishing Company, 1993.

2.1-3 C. S. Williams, "Designing Digital Filters", Englewood Cliffs, Prentice-Hall Inc. 1986, (ISBN 0-13-201856-X 01).

2.1-4 Rabiner, L., and Juang, B.-W., "Fundamentals of Speech Recognition", Englewood Cliffs, NJ, Prentice-Hall, 1993, (ISBN 0 13 015157 2).

2.1-5 Hermansky, H., "Perceptual linear predictive (PLP) analysis of speech", J. Acoust. Soc. Am., 87(4), April 1990, pp 1738-1752.

2.1-6 Hunt, M. J., and Lefèbvre, C., "A Comparison of Several Acoustic Representations for Speech Recognition with Degraded and Undegraded Speech", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", 1989, pp 262-265.

2.1-7 Davis, S. and Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-28, No. 4, August 1980, pp 357-366.

2.1-8 Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K., "Phoneme recognition using time-delay neural networks", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", 1988, pp 99-102.

2.1-9 Rabiner, L., Levinson, S., and Sondhi, M., "On the application of vector quantization and hidden Markov models to speaker-independent isolated word recognition", Bell Sys. Tech. Journal, 62, 1983, pp 1075-1105.

2.1-10 Forney, G. D., "The Viterbi Algorithm", Proc. IEEE, Vol. 61, 1973, pp 268-278.

2.1-11 Lowerre, B. T., "The Harpy Speech Recognition System", Doctoral Thesis, Carnegie-Mellon University, Pittsburgh, PA, 1976.

2.1-12 Bourlard, H. and Wellekens, C. J., "Links between Markov models and multi-layer perceptrons", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, 1990, pp 1167-1178.

2.1-13 Renals, S., Morgan, N., Bourland, H., Cohen, M., and Franco, H., "Connectionist Probability Estimators in HMM Speech Recognition", IEEE Trans. On Speech and Audio Processing, Vol. 2, No. 1, Part II, 1994.

2.1-14 McGuinness, "Effects of Feedback Modality in an Airborne Voice Communications Task", RAE Technical report TR 87072, December 1987.

2.1-15 Wickerhauser, M. V., "Adapted Wavelet Analysis form Theory to Software", A. K. Peters, 1994, Massachusetts.

2.1-16 Wesfreid, E., and Wickerhauser, M. V., "Adapted Local trigonometric Transforms and Speech Processing", IEEE Transactions on Signal Processing, Vol. 41, No. 12, 1993.

2.1-17 Pastor, D., "Diagnostic sur Signaux quasi-stationnaires par Décomposition en Ondelettes Orthonormales et Détection de Coefficients Significatifs", Thèse de l'Université de Rennes I, 1997.

2.1-18 Lemoine, C., " Recherche de traits acoustiques de la parole bruitée par Analyse Multi-Résolution ", Thèse de l'Université de Bordeaux I, 1998.

2.1-19 Howard, J. D., "Flight testing of the AFTI/F-16 voice interactive avionics system", in "Proc. Military Speech Technology", 1987, pp 76-82.

2.1-20 Pastor, D., and Gulli, C., "Improving Recognition Rate in Adverse Conditions by Detection and Noise Suppression", in "Proc. ESCA Workshop on Speech Recognition in Adverse Conditions", 1992, Cannes-Mandelieu.

2.1-21 Pastor, D., and Gulli, C., "DIVA 5 Dialogue Vocal pour Aéronef : Performance in Simulated Aircraft Cockpit Environments", Joint ESCA-NATO/RSG10 Tutorial and Workshop: Applications of Speech Technology, 1993, Lautrach.

2.1-22 Simpson, C. A., Coler, C. R., and Huff, E. M., "Human Factors of Voice I/O for Aircraft Cockpit Controls and Displays", in "Proc. Of the Workshop on Standardization for Speech I/O Technology", National Bureau of Standards, Gaithersburg, Md., March 1982.

2.1-23 Hoskins, J. W., "Voice I/O in the Space Shuttle", Speech Technology, Media Dimensions, New York, Vol. 2, No. 3, Aug./Sept. 1984, pp 13-18.

2.1-24 Castiglione, D., and Goldman, J., "Speech and the Space Station", Speech Technology, Media Dimensions, New York, Vol. 2, No. 3, Aug./Sept. 1984, pp 19-27.

2.1-25 Salazar, G., "Voice Recognition Makes its Debut on the NASA STS-41 Mission," Speech Technology, Feb/March 1991.

2.1-26 Salisbury, M., and Chilcote, J., "Investigating Voice I/O for the Airborne Warning and Control System (AWACS)," Speech Technology, Media Dimensions, New York, Vol. 5, No. 1, Oct/Nov 1989, pp 50-55.

2.1-27 Benson, P., and Vensko, G., "A Spoken Language Understanding System for Phraseology Training of Air Traffic Controllers," Speech Technology, Media Dimensions, New York, Vol. 5, No. 1, Oct./Nov. 1989, pp 64-69.

2.1-28 Sumby, W. H. and Pollock, I., "Visual contribution to speech intelligibility in noise", J. Acoust. Soc. Am., Vol. 26, 1954, pp 212-215.

2.1-29 Summerfield, Q., "Some preliminaries to a comprehensive account of audiovisual speech perception", in B. Dodd and R. Campbell, (Eds) "Hearing by Eye: the Psychology of Lip-reading", Lawrence Erlbaum Associates, London, 1987, pp 3-51.

2.1-30 Reisberg, D., "Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli", in B. Dodd and R. Campbell, (Eds) "Hearing by Eye: the Psychology of Lip-reading", Lawrence Erlbaum Associates, London, 1987, pp 97-113.

2.1-31 Massaro, D.W., "Speech perception by ear and eye", in B. Dodd and R. Campbell, (Eds) "Hearing by Eye: the Psychology of Lip-reading", Lawrence Erlbaum Associates, London, 1987, pp 33-83.

2.1-32 Finn, K. E. and Montgomery, A. A., "Automatic optically-based recognition of speech", Patt. Recogn. Lett., 8(3), 1988, pp 159-164.

2.1-33 Mase, K., and Pentland, A., "Automatic lipreading by computer", in "Proc. IEICE Image Understanding Symposium", Apr. 1989, pp 65-70.

2.1-34 Mase, K., and Pentland, A., "Lip reading: Automatic visual recognition of spoken words", Opt. Soc. Am. Topical Meeting on Machine Vision, June 1989, pp 1565-1570.

2.1-35 Goldschen, A., "Continuous Automatic Speech Recognition by Lipreading", PhD thesis, The George Washington University, Washington, DC, 1993.

2.1-36 Petajan, E. D., "Automatic lipreading to enhance speech recognition", in "Proc. IEEE Global Telecom. Conf.", Nov. 1984, pp 265-272.

2.1-37 Petajan, E. D., "Automatic Lipreading to Enhance Speech Recognition", PhD thesis, University of Illinois, 1984.

2.1-38 Brooke, N. M., and Petajan, E. D., "Seeing speech: Investigations into the synthesis and recognition of visible speech movements using automatic image processing and computer graphics", in "Proc. Int Conf. Speech Input and Output: Techniques and Applications", Science Education and Technology Division of the IEE, Mar. 1986, pp 104-109.

2.1-39 Sejnowski, T. J., Yuhas, B. P., Goldstein, M. H., and Jenkins, R. E., "Combining visual and acoustic speech signals with a neural network improves intelligibility", in D. S. Touretzky, (Ed) "Advances in Neural Information Processing", Morgan Kaufman, 1990.

2.1-40 Yuhas, B. P., Goldstein, M. H., and Sejnowski, T. J., "Integration of acoustic and visual speech signals using neural networks", IEEE Commun. Mag., Nov. 1989, pp 65-71.

2.1-41 Yuhas, B. P., Goldstein, M. H., Sejnowski, T. J., and Jenkins, R. E., "Neural network models of sensory integration for improved vowel recognition", Proc. IEEE, 78(10), Oct. 1990, pp 1658-1668.

2.1-42 Stork, D. G., Wolff, G., and Levine, E., "Neural network lipreading system for improved speech recognition", in Intl. Joint Conf. on Neural Networks, 1992, pp 285-295.

2.1-43 Silsbee, P. L., "Computer Lipreading for Improved Accuracy in Automatic Speech Recognition", PhD thesis, University of Texas at Austin, 1993.

2.1-44 Stork., D. G., and Hennecke, M. E., (Ed) "Speech Reading by Humans and Machines: Models, Systems, and Applications", Berlin, Germany, Springer-Verlag, 1996, (ISBN 0 540 61264 5).

2.1-45 "Proceedings of the Fourth International Conference on Spoken Language Systems", 3-6 Oct., 1996, Philidelphia, PA.

2.1-46 "Workshop on Audio-Visual Speech Processing", 26-27 Sept., 1997, Rhodes, Greece.

2.1-47 Silsbee, P.L., and Su, Q., "Audiovisual Sensory Integration Using Hidden Markov Models", in Stork., D. G., and Hennecke, M. E., (Ed) "Speech Reading by Humans and Machines: Models, Systems, and Applications", Berlin, Germany, Springer-Verlag, 1996, pp 331-349, (ISBN 0 540 61264 5).

2.1-48 Hennecke, M. E., Stork, D. G., and Prasad, K. V., "Visionary Speech: Looking Ahead to Practical Speech Reading Systems", in Stork., D. G., and Hennecke, M. E., (Ed) "Speech Reading by Humans and Machines: Models, Systems, and Applications", Berlin, Germany, Springer-Verlag, 1996, pp 408-423 (ISBN 0 540 61264 5).

2.1-49 Bregler, C., Omohundro, S. M., Shi, J., and Konig, Y., "Towards a Robust Speechreading Dialogue System", in Stork., D. G., and Hennecke, M. E., (Ed) "Speech Reading by Humans and Machines: Models, Systems, and Applications", Berlin, Germany, Springer-Verlag, 1996, pp 408-423, (ISBN 0 540 61264 5).

2.1-50 Scherer, K. R., "Voice, Stress and Emotion", in M. H. Appley, R. Trumbull (Eds), Dynamics of Stress, New York, 1986, pp 157-179.

2.1-51 Ephraim, Y., Malah, D., and Juang, B.-H., "On the application of hidden Markov models for enhancing noisy speech", IEEE Trans. Acoust., Speech, Signal Processing, 37(12), December 1989, pp 1846-1856.

2.1-52 Erell, A., and Weintraub, M., "Estimation using log-spectral distance criterion for noise-robust speech recognition," in "Proc. Intl. Conf. Acoust., Speech, Signal Processing", 1990, pp 853-856.

2.1-53 Varga, A. P., and Moore, R. K., "Hidden markov model decomposition of speech and noise", in "Proc. Intl. Conf. Acoust., Speech, Signal Processing", 1990, pp 845-848.

2.1-54 Acero, A., and Stern, R. M., "Environmental robustness in automatic speech recognition", in "Proc. Intl. Conf. Acoust., Speech, Signal Processing", 1990, pp 849-852.

2.1-55 Ephraim, Y., "A Bayesian estimation approach for speech enhancement using hidden Markov models", IEEE Trans. Acoust., Speech, Signal Process., April 1992, 40(4), pp 725-735.

2.1-56 Boll, S. F., "Speech enhancement in the 1980s: Noise suppression with pattern matching", in Furui, S., and Sondhi, M. M., (Eds), "Advances in Speech Signal Processing", M. Dekker, New York, 1992, pp 309-325.

2.1-57 Ephraim, Y., "Gain-adapted hidden Markov models for recognition of clean and noisy speech", IEEE Trans. Acoust., Speech, Signal Process., 40(4), April 1992, pp 725-735.

2.1-58 Tishby, N., "A dynamical systems approach to speech processing", in "Proc. Intl. Conf. Acoust., Speech, Signal Processing", 1990, pp 365-368.

2.1-59 Ghitza, O., "Auditory nerve representation as a basis for speech processing", in Furui, S., and Sondhi, M. M., (Eds), "Advances in Speech Signal Processing", M. Dekker, New York, 1992, pp 453-485.

2.1-60 Junqua, J.-C., Wakita, H., and Hermansky, H., "Evaluation and optimization of perceptually based ASR front end", IEEE Trans. Speech and Audio Process., l(l), January 1993, pp 39-48.

2.1-61 Woods, W. A., "Language processing for speech understanding," in Waibel A., and Lee, K.-F., (Eds), Readings in Speech Recognition, Morgan Kaufman, San Mateo, CA, 1990, pp 519-533.

2.1-62 Young, S. R., Hauptmann, A. C., Ward, W. H., Smith, E. T., and Werner, P., "High-level knowledge sources in usable speech recognition systems", in Waibel, A., and Lee, K.-F., (Eds), "Readings in Speech Recognition", Morgan Kaufmann, San Mateo, CA, 1990, pp 538-549.

2.1-63 Lee, K.-F., "Large-Vocabulary Speaker-Independent Continuous Speech Recognition: the SPHINX system", PhD thesis, Carnegie-Mellon, 1988.

2.1-64 Lee, K.-F., Hon, H.-W., and Reddy, R., "An overview of the SPHINX speech recognition system", IEEE Trans. Acoust., Speech, Signal Process., 38(l), January 1990, pp 35-45.

2.1-65 Lee, K.-F., "Context-dependent phonetic hidden Markov models for speaker-independent continuous speech recognition", IEEE Trans. Acoust., Speech, Signal Process., 38(4), April 1990, pp 599-609.

2.1-66 Waibel, A., "Prosody and Speech Recognition", PhD thesis, Carnegie-Mellon, 1986.

2.1-67 Waibel, A., "Prosodic knowledge sources for word hypothesization in a continuous speech recognition system", in Waibel, A., and Lee, K.-F., (Eds), "Readings in Speech Recognition", Morgan Kaufmann. San Mateo, CA, 1990, pp 534-537.

2.1-68 Gulli, C., Pastor, D., Lèger, A., Sandor, P. B., Clere, J. M., and Grateau, P. "G-load effects and efficient acoustic parameters for robust speaker recognition" in "Advanced Aircraft Interfaces: the Machine Side of the Man-Machine Interface." AGARD-CP-521 Munich, Germany, May 1992

2.1-69 Pallett, D., and Fiscus, J., "1996 Preliminary Broadcast News Tests", in "Proc. of DARPA Speech Recognition Workshop", Feb. 1997, Virginia.

2.1-70 Kudo, I., Nakama, T., Watanabe, T., and Kameyama, R., "Data collection of Japanese dialects and its influence into speech recognition", in "Proc. Int. Conf. On Spoken Language Systems", Oct. 1996, pp 308-311.

2.1-71 Hansen, J. H. L., and Arslan, L. M, "Foreign accent classification using source generator based prosodic features", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", May 1995, pp 836-839.

2.1-72 Teixeira, C., Trancoso, I., and Serralheiro, A., "Accent Classification", ", in "Proc. Int. Conf. On Spoken Language Systems", Oct. 1996, pp 577-580.

2.1-73 Kumpf, K., and King., R. W., "Automatic Accent Classification of Foreign Accent Australian English Speech", in "Proc. Int. Conf. On Spoken Language Systems", Oct. 1996, pp 4-7.

2.1-74 Arslan, L. M, and Hansen, J. H. L., "Improved HMM training and scoring strategies with application to accent classification", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", May 1996, pp 598-601.

2.1-75 Arslan, L. M, and Hansen, J. H. L., "Language Accent Classification in American English", Speech Communication, Vol. 18(4), pp 353-367, June/July 1996.

2.1-76 Humphries, J. J., Woodland, P. C., and Pearce, D., "Using accent-specific pronunciation modelling for robust speech recognition", in "Proc. Int. Conf. On Spoken Language Systems", Oct. 1996, pp 623-626.

2.1-77 Diakoloukas, V., Digalakis, V., Neumeyer, L., and Kaja, J., "Development of a dialect-specific speech recognizers using adaptation methods", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", April 1997, pp 1455-1458.

2.1-78 Lee, C., Lin, C.-H., and Juang, B.-H., "A study on speaker adaptation of the parameters of continuous density hidden Markov models", IEEE Trans., 39(4), April 1994, pp 806-814.

2.1-79 Neumyer, L. and Wientraub, M., "Robust speech recognition in noise using adaptation and mapping techniques", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", 1995, pp 141-144.

2.1-80 Acero, A., and Stern, R. M., "Robust speech recognition by normalization of the acoustic space", in "Proc. Int. Conf. on Acoust., Speech, and Signal Processing", April 1991, pp 893-896.

2.1-81 Hermansky, H., personal communications, 1995.

2.1-82 Hirsch, H., Meyer, P., and Ruehl, H. W., "Improved speech recognition using high-pass filtering of subband envelopes", in "Proc. of Eurospeech 1991", September 1991, pp 413-416.

2.1-83 Murveit, H., Butzberger, J., and Weintraub, M., "Reduced channel dependence for speech recognition", in "Proc. of the DARPA Speech and Natural Language Workshop", Harriman, NY, February 1992, pp 280-284.

2.1-84 Stanton, B. J., "Robust recognition of loud and Lombard speech in the fighter cockpit environment", Doctoral Thesis, Purdue University, West Lafayette, IN, 1988.

2.1-85 Rajasekaran, P. K. and Doddington, G., "Robust speech recognition: initial results and progress", in "Proc. of the DARPA Speech Recognition Workshop", Palo Alto, CA, February 1986, pp 73-80.

2.2. HEAD-BASED CONTROL

2.2.1. INTENTION OF THE TECHNOLOGY

The principal objective for "head tracking" devices is to measure head translation and orientation with respect to an airframe (or workstation), either for direct use as a pointing mechanism, to slave an airframe mounted sensor, or to provide necessary data to head mounted display systems or eye tracking systems.

2.2.1.1. Relevance

Head position measurement can be used for explicit control functions, in which aircrew purposely move their head to affect a control input, or for implicit functions in which natural head motions are used to slave an external sensor or to appropriately place elements on the image created by a head mounted display.

A head tracker can be used explicitly as a pointing device by requiring aircrew to sight target objects through a head mounted reticule. This technique has seen operational use as a means of designating external targets for weapons delivery, and is discussed later on in section 2.2.3. Head position pointing could similarly be used for switch selection and cursor control tasks within the cockpit. Although gaze measurement, as discussed in section 2.3, might provide more potential speed for target designation or other pointing tasks, head tracking technology is currently more mature than eye tracking technology in terms of flight worthiness and dependability.

Head tracking can also be used to slave airframe mounted sensors (e.g., a night vision sensor) to helmet position. The sensor images can then be displayed on a helmet mounted display. This is an example of implicit control since the pilot simply moves his head naturally, but is always able to see the appropriate sensor image in his central gaze region. This application has also seen operational use and is discussed in section 2.2.3.

The example of a head slaved sensor is a case in which a head mounted display image is made to appear stable with respect to the outside environment. Head position measurement is also required if a head mounted display image must appear to be stabile with respect to the cockpit. Future use of virtual environments, for example, may require that images of virtual controls created by a helmet mounted display appear to be fixed to the airframe.

In the future, eye line of gaze may be used to designate targets, or to interact with objects and switches in virtual environments (see section 2.3). If eye position is measured with respect to the headgear, as it probably will be, head position and orientation measurement is required to determine line of gaze with respect to the airframe. In this case, a head tracker must be an integral part of the line of gaze measurement system.

Finally, head position measurement may prove to be a useful component for gesture recognition, as discussed in section 2.4.

Designation of distant external targets, either using head control alone or as part of an eye gaze measurement system, may only require measurement of head orientation (3 rotational degrees of freedom) since the parallax effect of motion within the cockpit is negligible. Pointing tasks within the cockpit as well as some head mounted display stabilisation tasks require measurement of head position as well as orientation (all 6 degrees of freedom).

2.2.1.2. Human capabilities and Limitations

The head can be thought of as an inverted pendulum resting on the cervical vertebrae, and supported and controlled by numerous large muscle pairs. The range of possible voluntary head motion with respect to the trunk, at the joint of the neck, was measured in male civilians by Glanvilee and Kreezer [2.2-1] and is reported by Hertzberg [2.2-2]. Average values were 60° and 61° respectively for ventral and dorsal flexion (chin up, chin down), ± 41° degrees for lateral flexion (right or left ear towards shoulder), and ± 79° of rotation about the spinal column axis. Standard deviations were close to 20% of all of the average values except dorsal flexion (chin up) which had a standard deviation of 44%. Durlach and Mavor [2.2-3] report typical peak velocities for voluntary head motion of about 600°/sec in yaw rotation, and about 300°/sec in pitch, with virtually all frequency domain energy below 15 Hz.

Typical reaction to the appearance of a non predictable visual target is an eye saccade, followed by a head motion towards the target after an additional delay (beyond initiation of the eye saccade) of 30-50 msec. If the target appearance time and location is predictable, an anticipatory head motion typically precedes eye motion by up to several hundred msec [2.2-4, 2.2-5, 2.2-6, 2.2-7].

Barnes and Sommerville [2.2-8] found that target acquisition through a head mounted sight typically required 0.8-1.5 seconds to bring the aim point to within about 2.5° of the target, and 2-4 seconds to come within 0.3° of the target. It should be noted that while turning the head toward a target is a natural action, neither fine positioning of the head nor maintaining rigid head positions for extended periods are at all natural.

A comprehensive review of human ability to track and aim with head-mounted sights has been conducted by Wells and Griffin [2.2-9]. The authors investigated, under static laboratory conditions, the effect of operationally related variables such as off-boresight target angle, reticule size and shape, helmet weight, etc., and concluded that these variables had a minor impact on performance.

Sandor and Leger [2.2-10] studied head movements, under laboratory conditions, while subjects performed a visuo-manual compensatory tracking task with a reduced field of view. When the field of view was sufficiently reduced not only was tracking performance impaired, but basic eye-head coordination mechanisms were altered. Head movement alone was used to shift gaze, thereby creating a greater than normal need for head stability in order to perform the tracking task well.

In dynamic environments, especially aerospace environments, the effect of mechanical forces on head motor control is particularly important. Viviani and Berthoz [2.2-11] studied the dynamics of the head-neck system in response to small perturbations. They showed that, when a subject was

instructed to voluntarily resist a sinusoidally applied force, sinusoidal head displacement was significantly distorted for frequencies below 2 hertz. Above this frequency the response appeared to be almost linear, suggesting that the system behaves as a quasi-linear second order system with two degrees of freedom. This demonstrates that static force analysis is not sufficient to predict head motor control behaviour in the presence of external forces and that consideration of dynamics is essential.

The transmission of vibration from the seat to the head of an operator has been studied extensively [2.2-12, 2.2-13, 2.2-14. 2.2-15], and although the seat design, harness tension, posture and body type exert strong influences, the predominant disturbance is an involuntary nodding of the head due to vertical (heave) seat motion. Sideways (sway) and fore/aft (shunt) motions have significantly less effect. Head pitching can be controlled voluntarily if the excitation is below about 0.5 Hz, while vibration above about 10 Hz is damped by the trunk, so that most people are affected strongly by vibration between these frequencies, particularly in the 3 to 6 Hz region; a little above the frequency of jolts transferred to the head of a runner. Thus head pointing is disturbed significantly by whole-body vibration [2.2-9, 2.2-16].

In modern aerospace environments, especially fast jets, there is often significant Gz loading. It has been reported that the head begins to feel noticeably heavy at 2-3 Gz, and cannot be lifted at all after 8 Gz [2.2-17]. From the mid 1980s to 1990, centrifuge studies were conducted at the French Flight Test Center (CEV), as part of the "Rafale" program, to assess the effect of Gz loads on head aiming accuracy. [2.2-18, 2.2-19, 2.2-20, 2.2-21]. The main findings were that accuracy of aiming at fixed targets and tracking performance were moderately affected by constant Gz loads (up to 5 Gz). RMS errors were found to be 0.2° under normal gravity, increasing to 0.8°-1° at 5 Gz. Aiming accuracy was shown to be strongly affected by Gz onset rate, with RMS errors averaging up to 1.5° for an onset rate of 1 Gz/sec. In some cases, maximum errors during the onset phase were larger than 5°. These findings are consistent with the results of basic studies showing that head motor control adapts quite well to static constraints [2.2-9] but does not perform nearly as well in dynamically changing conditions.

Incidental observations made during tracking experiments have indicated some important biomechanical effects of Gz [2.2-22]. With their heads resting against the seat head rest, subjects tried to continuously track a target moving from a display center to the periphery (50° azimuth and elevation). Under conditions of 5 Gz, several subjects could not rotate their head farther than about 20° in azimuth and 40° in elevation. This effect was never observed during ballistic head movements towards peripherally positioned targets or when the target was moving from the periphery to centre. It can be hypothesised that the effect is explained by changes in head centre of gravity location with respect to neck pivot points. Head worn mass and centre of gravity location are also related to fatigue and safety issues, and should be considered with regard to head mobility under Gz. For example, things that appear to alleviate fatigue under static conditions, such as artificially increasing head stability, may have detrimental effects on head mobility in dynamic environments.

For several years now; helmet mounted systems including head trackers have been flight tested in various countries. Some have already been fielded. Recent reports from flight tests in France confirm laboratory results, in particular in the area of varying accelerations.

2.2.2. OVERVIEW OF APPROACHES

The predominant techniques for measuring head position and orientation can be classified as mechanical, inertial, acoustic, optical, and magnetic. Mechanical, optical and magnetic head tracking techniques have already seen some operational use in military aircraft. In recent years magnetic systems have probably seen the widest use and can be considered a relatively mature technology for the aerospace environment. Although some specific implementations have been designed to measure only head orientation, all categories of system can theoretically measure all 6 degrees of freedom.

Translation measurements (3 degrees of freedom) specify the location of a fixed point on the head gear with respect to a fixed origin in the airframe. Translation is typically specified in Cartesian coordinates, but can also be specified in polar coordinates. Orientation measurements (3 degrees of freedom) specify the orientation of a coordinate frame fixed to the head gear relative to an airframe fixed coordinate frame. Rotation is typically specified as 3 Euler angles, a 9 element rotation matrix, or a set of 4 quarternions (see Appendix E).

Head tracker performance is often described in terms of some of the following parameters, usually specified separately for translation and orientation measures. *Accuracy* is the expected difference between measured position and true position. *Precision* (repeatability) is the expected difference in repeated measurements of the same true position. *Resolution* is the smallest change in true position that can be reported by the device. *Range* is the maximum excursion from some specified nominal position over which valid measurements can be made. Orientation range is usually specified in terms of the three Euler angles, and translation range is usually specified as a three dimensional region of space ("motion box"). *Update rate* is the frequency with which data samples are measured and reported, usually reported as "samples/second". *Transport delay* is the amount of time that it takes data to travel through the system and become available for use. *Latency* (or *throughput* as defined by Kocian and Task [2.2-23] usually refers to the amount of time required to accurately reflect a change in the quantity being measured. It is influenced by pure transport delay and also by dynamic operators (for example, a low pass filter) in the signal path. A more detailed discussion of performance parameters can be found in Kocian and Task [2.2-23]. *Bandwidth* is the range of sinusoidal input frequencies that can be processed by the system without significant attenuation or distortion.

2.2.2.1. Mechanical Tracking

2.2.2.1.1. Mechanical Tracking - Technique

Mechanical head trackers, sometimes referred to as goneometers, work by mechanically coupling head gear to the environment (e.g. airframe) through a set of linkages connected by flexible joints. The position of each joint is measured by a transducer, and the set of joint positions is used to calculate head gear position and orientation in six

degrees of freedom. Transducers are typically optical encoders, potentiometers, strain gauges, or some combination of these.

There are a very small number of commercially available mechanical devices which are specifically designed to track head gear position and orientation. One such device, for example, consists of two long arms connected to each other with a single joint (one degree of freedom), and fastened through multiple joints to a fixed base at one end and to the head gear at the other. Joint angles are measured with potentiometers. Other commercially available systems were designed as scribing tools to trace and digitize the shape of 3D surfaces; but might be adapted for head tracking. Many "one of a kind" goneometers have been built for use in research and simulation laboratories. One such device, developed for use with a flight simulator [2.2-24] is sketched in Figure 2.2-1.

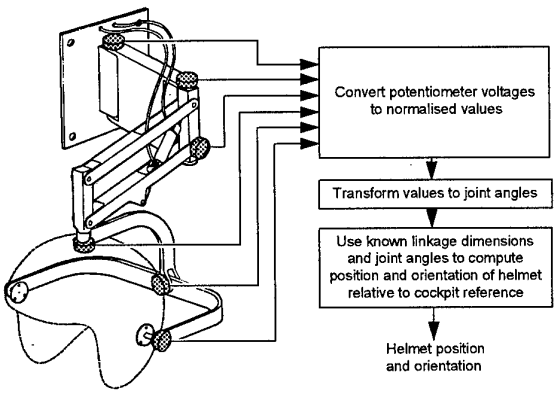


Figure 2.2-1 Sketch of mechanical head tracker built for use in a flight simulator, redrawn from [2.2-24].

Some custom mechanical systems have been flight tested in various countries, especially on helicopters, and are usually designed to provide only azimuth and elevation degrees of freedom. For example, such a system has been used on the Cobra helicopter for many years. The system used on the Cobra [2.2-25] consists of an overhead slider mechanism allowing a rod to slide fore and aft just above the pilots head. The rod is attached to the slide track with a universal joint and also has a universal joint on the other end which can be attached, via a nipple-shaped magnet, to a mating receptacle on the pilot's helmet. The magnetic helmet attachment mechanism allows for very quick disconnect. The universal joint angles are measured with AC resolvers. Analog outputs from the resolvers are input to a head tracker electronics unit which computes the azimuth and elevation angle of the pilot's helmet (2 degrees of freedom), and sends a corresponding command signal to a rotating gun mount, or to a wire guided missile system.

2.2.2.1.2. Mechanical Tracking - Performance

Mechanical trackers can have relatively low cost, and are capable of good accuracy, high update rate, reasonable range for a seated user, and very good dependability. Inexpensive systems should easily achieve accuracy of 0.2 inches translation and 0.2° orientation, updates rates above 500 samples/sec, and working volumes of 2-3 ft in diameter. Some commercial systems designed as scribing tools quote position accuracy of 0.01 inch.

2.2.2.1.3. Mechanical Tracking - Practical Problems

For in-flight use, there are some profound practical problems. The very presence of a mechanical link between a pilot's head gear and the airframe poses an enormous problem for ejection safety. In aircraft that do not use ejection, any helmet latching mechanism must be designed to allow very rapid de-latching and egress. The joints and linkages take up valuable cockpit space, are subject to mechanical damage and are affected by inertial forces. Even for ground based applications, there may be some resistance on the part of users to being mechanically linked to the environment.

2.2.2.1.4. Mechanical Tracking - Prognosis

Due to the practical problems cited above, and in spite of excellent performance parameters, future in-flight use of mechanical head trackers is likely to be restricted to helicopter, transport, or ground-based applications, and then only when low cost is important. Mechanical trackers will probably continue to be extremely useful as low cost research and development tools. As previously noted, they have seen operational, in-flight use in the past.

2.2.2.2. Inertial Tracking

2.2.2.2.1. Inertial Tracking - Technique

Inertial sensors are available which can measure angular velocity and specific force (the vector sum of gravity and acceleration forces) with respect to an inertially stable reference frame. Historically, the first practical angular velocity measurement device was a rate gyroscope, which in its simplest conceptual form is a single degree of freedom spinning wheel gyroscope restrained, by a spring, from rotating about an axis perpendicular to the gyro spin axis. If the device is rotated about an axis perpendicular to both the gyro spin axis and spring restraint axis, the spring is stretched or compressed by an amount proportional to the rotational velocity. A package of three orthogonal rate gyroscopes theoretically measures the total angular velocity vector. The accelerometer (specific force measurement device) in its simplest conceptual form is a mass attached to a spring and constrained to move only along the axis of the spring. In this case the spring is stretched or compressed proportionately to gravity or acceleration forces along the spring axis. Three orthogonal accelerometers theoretically measure the total specific force vector.

Modern practical versions of the these inertial components often employ some combination of micro machined tuning forks, miniature spinning wheels, ring lasers, and piezoelectric elements, and sometimes employ a more complex set of physical principles to achieve improved performance in small packages. The measured quantities, however, remain basically the same: angular velocity and specific force.

Methods for position and orientation tracking with such instruments have been developed for inertial navigation and the same principles can be applied to tracking a person's head gear. If an initial orientation is known, angular velocity can be integrated to continually estimate orientation angle. Once orientation with respect to gravity is known, gravity can be subtracted from specific force data to yield acceleration with

respect to the gravitational field. If an initial position and velocity are known, acceleration can then be integrated to continually estimate current position and orientation.

Transient errors in the angular velocity or acceleration measurements accumulate in the integrated orientation and position estimates. Even if the inertial components are quite accurate this "dead reckoning" technique requires periodic independent measures of position and orientation to remove accumulated drift. The rate of drift, and consequently the frequency with which it must be corrected, depend on the accuracy of the sensor measurements and of the integration process.

Sophisticated aircraft inertial navigation systems include optimal estimation algorithms which use knowledge of measurement noise characteristics and error mechanisms, and expected statistical characteristics of the actual motion, to make the best possible estimates of state. Satellite navigation data can be used to periodically reset position estimates. Attitude measurement systems can also take advantage of knowledge that, over the long term, the average specific force vector will be aligned with gravity.

Head tracking with inertial sensors requires that a package of angular rate sensors and accelerometers be mounted to the user's head gear. The resulting size and weight constraints preclude use of the instrument packages that have been developed for aircraft navigation and attitude detection, but small angular rate sensor and accelerometer components are available. Inertial sensors measure motion with respect to an inertially stable reference frame; so in order to measure head motion with respect to an aircraft cockpit, information from an inertial package that is fixed to the airframe must be subtracted from measurements made by the head mounted package.

The requirement for frequent drift correction constrains inertial head tracking to use in conjunction with other head tracking techniques. There is currently a commercially available system that uses a combination of acoustic and inertial sensors to measure head gear position and orientation. An early version of this device is described in Foxlin and Durlach [2.2-26]. The head mounted inertial package measures approximately 3.5 cm x 3 cm x 3 cm. The device was not, however, intended for use on an aircraft and the current system makes no provision for subtracting vehicle motion. Inertial sensors, particularly angular rate sensors, have been used quite successfully to add high frequency (lead) information to systems employing other head tracking techniques. For example, Emura and Tachi [2.2-27] describe an optimal estimation technique for combining inertial angular rate information with magnetic head tracker data.

2.2.2.2.2. Inertial Tracking - Performance

Inertial sensors provide high bandwidth angular velocity and acceleration information, and can provide position and orientation information with very high resolution. Foxlin and Durlach [2.2-26], for example, report 0.008° orientation resolution with inertial components. Steady state accuracy of position and orientation data is very poor due to accumulated integration errors (drift). With a sensor package that is of practical size and weight for head mounting, drifts of at least several degrees/minute and several cm/minute would not be unexpected.

Depending on the specific implementation, angular rate and specific force accuracy for head mounted packages are likely to be in the range of 0.1 to 1 °/sec and 0.002 to 0.2 m/sec^2 respectively. Update rates of at least 500 samples per second should be possible.

2.2.2.2.3. Inertial Tracking - Practical problems

Drift of steady state position and orientation values is the overwhelming problem for inertially based head tracking.

2.2.2.2.4. Inertial Tracking - Prognosis

Inertial head tracking alone is unlikely to be practical due to undependable steady state performance, however it may be very useful for providing high frequency information in conjunction with other systems that have dependable steady state performance. The use of inertial sensors coupled with earth magnetic field sensing and with modern navigation devices such as GPS are envisioned for military ground applications (Land Warrior).

2.2.2.3. Acoustic Tracking

2.2.2.3.1. Acoustic Tracking - Technique

Acoustic trackers use a triangulation technique that is usually based on sound propagation time. The ultrasonic frequency range is generally used so as not to be audible to people.

Assuming that the speed of sound is known, the delay between sound emission by a speaker, and detection by a microphone yields the distance between speaker and microphone. Note that this assumption can be compromised by changes in the speed of sound due to temperature changes or other atmospheric changes. Distance values from 3 known fixed receivers (microphones) to a moving speaker allows the emitter (speaker) position to be triangulated. The emitter is usually the moving component since a single emission from one speaker can easily be received by multiple microphones without confusion.

Line of sight must always be maintained between the emitters and receivers since it is assumed that sound can follow a straight trajectory between emitter and receiver.

If at least 3 such speakers are fastened in known positions on a helmet, the helmet position and orientation can be unambiguously computed.

A small number of commercially available systems have been designed primarily for use as 3D computer input devices. A device was made in the 1980s to acoustically detect pilot head orientation (3 rotational degrees of freedom) for weapon aiming application, but is no longer available. A commercially available device mentioned in the previous section on inertial tracking [2.2-26], combines acoustic steady state measures with higher bandwidth inertial measures to implement a head tracking device. The resulting system is intended to have update and throughput rates as well as resolution characteristics (ability to measure small changes) that are associated with inertial systems, while maintaining the steady state performance characteristics of acoustic trackers.

It is also possible to detect motion of an emitter with respect to a receiver by measuring phase changes between a signal and reference sound source [2.2-28]. This has the same inherent problem as inertial sensing in that no steady state measurement is made; rather, a velocity measure must be integrated. Applewhite [2.2-29] has proposed a variation of

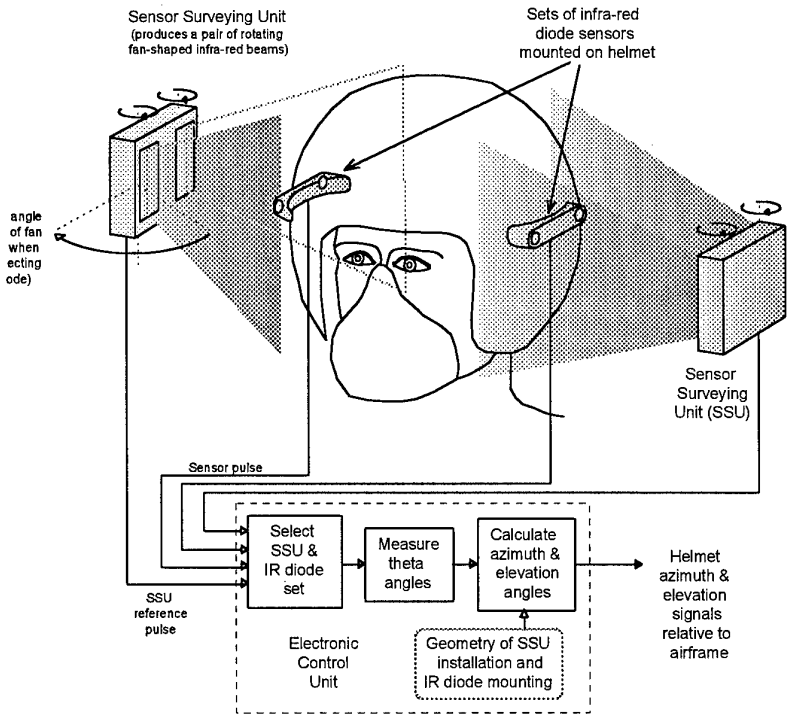


Figure 2.2-2. Schematic summarising the Honeywell MOTVAS optical head tracker

this phase coherence method to enable steady state measurement.

2.2.2.3.2. Acoustic Tracking - Performance

For most acoustic 6 degree of freedom tracker systems update rate is limited, by sound wave travel time and the need for multiple emitters, to the order of 30 samples/sec or less. If the phase coherence method is successfully used, much higher bandwidth will become possible. Current acoustic systems generally achieve static accuracies on the order of 0.5° rotation, and 5 mm translation.

Ultrasonic microphones have a receptive field typically on the order of a 45° radius cone. Assuming the microphones are fixed to the environment, as is usually the case, the tracking system motion box is constrained to the area where receptive fields from all microphones overlap. As the distance from speakers to microphones increase, noise and echo problems become more severe, and update rates become more restricted (due to increased time of flight). Typical maximum range for acoustic systems is probably on the order of 5 m, with best performance probably within 1 m.

2.2.2.3.3. Acoustic Tracking - Practical problems

Acoustic trackers require line of sight between emitters and receivers, are easily influenced by temperature gradients and air currents, and are subject to interference from echoes and other acoustic sources, especially in the noisy environment of military aviation.

2.2.2.3.4. Acoustic Tracking - Prognosis

Currently available acoustic tracking devices are not as accurate or dependable as the state of the art magnetic or optical tracking devices, and militarized versions are not currently available. Acoustic devices do not suffer from metal and electro-magnetic interference as do magnetic systems, or from sunlight interference as do optical trackers; but the problems listed above are at least as severe. Future development of acoustic technologies may solve or reduce the practical problems, but at present both magnetic and optical technologies are significantly more mature and are more likely to find practical use in airborne environments.

2.2.2.4. Optical Tracking

2.2.2.4.1. Optical Tracking - Technique

Over the past 35 years engineers have developed a variety of optical helmet tracking systems in an attempt to attain a satisfactory balance between measurement accuracy and reliability in the cockpit environment. Although several have exploited phenomena such as interferometry and pattern recognition [2.2-30], the most successful have been based upon triangulation. These invariably use near infra-red light, which is unnoticeable to the user and for which a variety of commercial emitters and receivers are available, and they all measure a set of angles between cockpit- and helmet-mounted devices. They differ by employing alternative devices, and in some the emitters are fixed in the cockpit while in others they are on the helmet. Their sensitivity to artifacts, particularly those due to incident sunlight, also depends strongly on the chosen sensor. An outline of the principles of operation of two airborne systems provides a nice illustration of the developer's technical ingenuity.

The Honeywell MOVTAS (Modified Visual Target Acquisition Set), shown schematically in Figure.2.2-2, was devised in the late 60's and has been installed in a variety of aircraft. It is best known as the helmet tracker employed in the IHADSS (Integrated Helmet and Designating Sub-System) for the AH-64 Apache helicopter, in current US

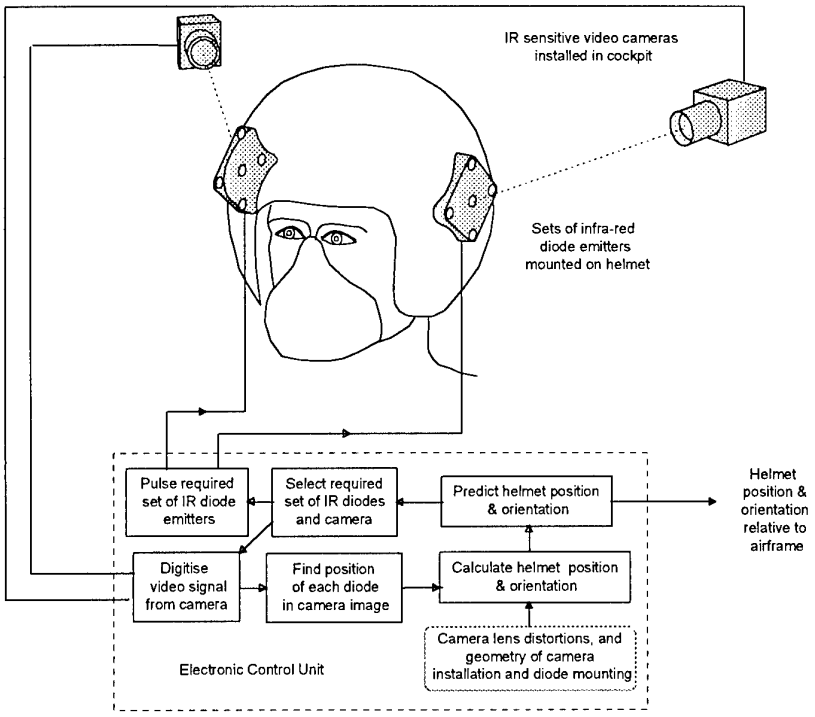


Figure 2.2-3. Schematic summarising a modern optical head tracker

Army service. As illustrated, a helmet-mounted infra-red sensing diode produces a short electrical pulse when swept by a fan-shaped beam from a sensor surveying unit (SSU) mounted in the cockpit, in a manner analogous to a sailor observing the flash of a lighthouse. As the beam rotates rapidly at a constant angular rate, the interval between a pulse produced by the helmet-mounted diode and a reference pulse produced by the beam rotating mechanism is proportional to the beam angle to the diode. The diodes are paired, and a pair is "surveyed" by both beams in a SSU to give four quasi-simultaneous beam angle measurements. Given knowledge of the installation dimensions, the electronic unit solves the trigonometric equations to calculate the helmet pointing direction, which is output each computational cycle as the helmet azimuth and an elevation angle. Several sets of diodes and SSUs are normally used to extend the range of measurement and the head box.

A more modern approach is illustrated in Figure 2.2-3. Here, a cluster of LED emitters on the helmet is imaged by a cockpit-mounted camera. An electronic unit, based on digital signal processing (DSP) chips, finds the position of each diode in the 2-dimensional camera image and, knowing the installation geometry and the distortion introduced by the camera optics, calculates both the position and the orientation of the helmet. The update rate of systems employing video cameras as imaging sensors is usually limited by the frame rate of the video signal to either 50 or 60 Hz, although fast frame cameras can be employed to increase the measurement frequency. Measurement delay can be virtually eliminated by motion prediction algorithms, and sensitivity to sunlight can be reduced significantly by only opening the camera electronic shutter during the brief fraction of the frame period when the diodes are pulsed.

Some systems use lateral effect photo-sensitive detectors (LEPSD) instead of video sensors [2.2-30] to increase the measurement update rate and enable sequential pulsing of individual diodes to remove any uncertainty in their identity and improve signal detectability. It is, however, essential to filter the incident light to exclude all but the IR source waveband to prevent sunlight from saturating the detector, but it is possible to compensate for the in-band sunlight by sampling the LEPSD output when all the diodes are momentarily inactive.

As with the MOVTAS system, the range of measurements and the allowable head box of the imaging techniques are invariably extended using several clusters of emitters and several cameras. The allowable range of head positions has been taken further in a ground-based laboratory where the user can walk around a room in which the ceiling is studded with clusters of IR emitters [2.2-31].

2.2.2.4.2. Optical Tracking - Performance

The configuration of sensors and emitters must be designed for each application, for instance an aircraft type, to match performance requirements with geometric constraints. Once the components have been installed in an individual aircraft cockpit to comply with the design, residual alignment errors can be corrected within the electronic control unit. Routine in-service calibration is unnecessary.

With pairs of SSUs and diode sets, MOVTAS measures the two primary helmet-pointing angles over a range of ±180° azimuth and ±70° elevation to an accuracy of about 0.5° (CEP), updated at 30 Hz with a delay of about 30 msec.

The performance of the modern optical tracking systems is considerably improved, primarily because all six degrees of motion are measured. The range of measured positions and angles varies with the installation, but ±180° in azimuth, ±60° in elevation and ±40° in roll combined with ±150 mm motion

in all three directions is practical. Static errors would be less than about 0.2° in angle and about 1 mm in position.

2.2.2.4.3. Optical Tracking - Practical Problems

The helmet-mounted and cockpit-mounted units must be installed where they give the required range of measurement and an adequate head motion envelope without intruding on the pilot's view through the canopy. The sensors should also be shielded from direct sunlight, and the canopy should not reflect either the sun or the IR emissions into the sensor field.

The usual configuration for imaging trackers is to install a pair of cameras as high and wide as possible behind the seat, angled downwards and inwards to point slightly below the normal helmet center. Here they are least susceptible to reflections, and the chance of obstruction by the pilot's arm or hand is negligible. The diodes may be attached to the rear and sides of the helmet, and the emission diffused to illuminate the sensors throughout the head motion range. However, as it is necessary to measure and maintain their relative positions to an accuracy of about 0.1 mm to achieve the advertised performance, the useful spread of diode sites may be limited by headgear flexure. It should, however, be borne in mind that the optical tracker has the advantage of some flexibility and additional emitters and cameras can be introduced to improve measurement performance, albeit at increased cost and complexity. For instance, a third camera pointing backwards from the cockpit front may be necessary to assure full azimuth coverage.

At night, the mixture of emitted, reflected and scattered IR from the SSUs makes MOVTAS incompatible with the use of night vision goggles. A similar intensifier overloading can occur with the later optical trackers, particularly when helmet-mounted diode emissions are reflected from the canopy. There is also some concern that IR emission from the cockpit could make military aircraft more readily detected by external surveillance systems.

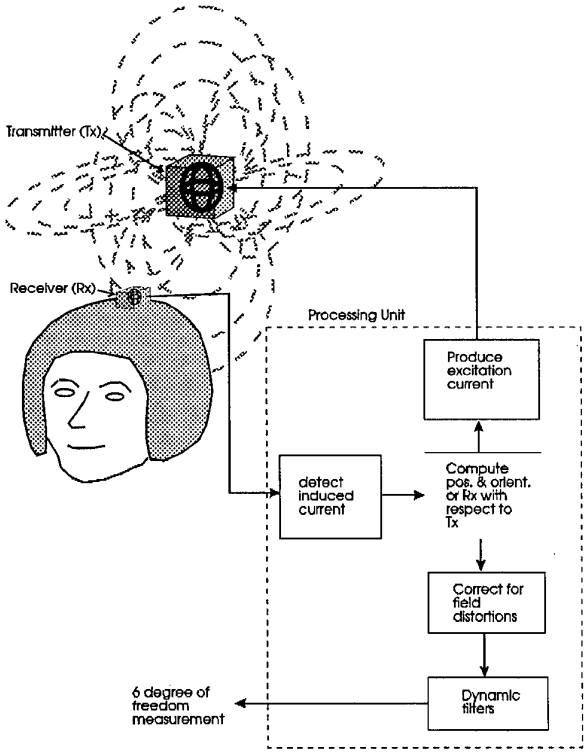


Figure 2.2-4. A generic electro-magnetic head tracking system

2.2.2.4.4. Optical Tracking - Prognosis

Although optical trackers offer good performance and require no calibration or alignment in service, they may be susceptible to strong sunlight during daytime and at night they may interfere with other cockpit systems which utilize the IR spectrum. Given that electro-magnetic tracking systems achieve comparable performance with none of these attendant drawbacks, and at similar cost, optical techniques are unlikely to be preferred.

It is possible that a simple optical tracker, working around a small cone of angles centred on the boresight, could be installed to complement an electro-magnetic system. The optical tracker could have the very high accuracy for delivering boresighted weapons, and it could alleviate the need for pre-take-off harmonization of the electro-magnetic system. Cross-checking would also ensure that the helmet tracker of a visually-coupled system was unlikely to produce erroneous, and potentially disorienting, measurements.

2.2.2.5. Magnetic Tracking

2.2.2.5.1. Magnetic Tracking - Technique

Magnetic trackers create magnetic fields of known orientation and measure the current induced in sensor (receiver) coils that are fixed to the object being tracked.

As shown schematically in Figure 2.2-4, a set of stationary antennae in the form of 3 orthogonally oriented coils mounted to the environment (e.g., airframe) are sequentially excited with electric current, sequentially producing electro magnetic fields with mutually orthogonal polarization. This set of antennae is usually referred to as the transmitter or source.

A smaller set of orthogonal oriented antennae, usually referred to as the sensor or receiver, are mounted to the object being tracked (e.g., aircrew headgear). The current induced in each of the 3 sensor antennae is measured during the field produced by each of the 3 transmitter antennae. The 9 sensor responses are processed to compute position of the sensor with respect to the transmitter in 6 degrees of freedom [2.2-32, 2.2-33].

The transmitter is typically housed in a cube shaped enclosure, ranging from 5.5 to 10 cm on each side. The sensor is typically housed in a much smaller enclosure, typically 1.5 to 2.5 cm on each side.

Two categories of magnetic system are available: those using an AC coupled technique and those using a DC technique. AC type systems excite each transmitter antenna with a sinusoid and can take advantage of AC coupling techniques to eliminate the effect of static fields in the environment. AC systems are very susceptible, however, to error due to the presence of conductive metal in the environment. The errors are due to eddy currents induced in the conductive metal by changing fields.

DC systems excite each transmitter antenna with a DC current pulse. Sensor antennae are sampled when the transmitter is dormant, as well as during the time each transmitter antenna is excited, so that components of the Earth's magnetic field

can be subtracted. When run with update rates in the region of 100 Hz, DC systems are far less sensitive to the presence of conductive metals than are AC systems. The eddy currents produced by field changes die out at an exponential rate proportional to the metal conductivity. As update rates increase and there is less time during each transmitter antenna pulse to wait for eddy currents to die away, DC systems become more susceptible to eddy current interference [2.2-34, 2.2-35].

This is now a relatively mature technology, and magnetic tracking devices of both AC and DC type are readily available in both commercial and militarized versions.

2.2.2.5.2. Magnetic Tracking - Performance

In a benign environment (no large metal objects or electro-magnetic interference problems), commercial type systems typically offer accuracy ranging from 0.75 to 2.5 mm translation, and 0.15-0.5° orientation. Accuracy is usually best when sensor and transmitter are very close, and tends to decrease as they separate. The allowable motion box is typically on the order of a 1 meter hemisphere for best performance. Update rate typically ranges from 60 -120 Hz, and latency ranges from 4-150 msec with a typical value of about 40 msec depending on the type of system and amount of filtering used.

Depending on the environment, varying amounts of filtering may be needed to reduce noise in the measurement. The filters used are usually dynamic filters with properties that are related to motion rates, and this makes latency determination very complex. A comparison of latencies in some commercially available systems can be found in [2.2-36].

A promising approach to reducing system lag, as described by Emura and Tachi [2.2-27] (and previously mentioned in the section describing inertial trackers), is to augment the magnetic system data with information from inertial sensors. The magnetic system provides accurate low frequency information, while angular velocity sensors can provide very good high frequency information.

It has been reported that a militarised magnetic tracker, developed to have a high degree of metal tolerance, has achieved angular accuracies of 0.1° RMS, within mapped areas, even in environments containing a great deal of interfering metal. This performance has been achieved, for example, in an OH-58 helicopter cockpit for sensor motion within an 18" x 12" x 7" motion box [2.2-37].

2.2.2.5.3. Magnetic Tracking - Practical Problems

Metal objects produce errors whose magnitude depends on proximity to the magnetic components as well as size and composition of the metal object. It has long been possible to compensate for effect of stationary metal, but determination of compensation equation parameters is an elaborate procedure requiring placement of the sensor in many precisely known positions with a non metallic jig. The results are then valid only for one precisely defined physical environment. Such procedures, referred to as cockpit mapping, usually take several days to be completed with an acceptable accuracy. Although results are acceptable, the time and effort required for the mapping process constitute a significant problem. Transfer of mapping data from one aircraft to another of same type is also a serious practical problem, closely linked to manufacturing tolerances.

A more difficult problem has been posed by metal objects attached to the aircrew head gear, and subject to repositioning as helmet mounted systems are reconfigured for different tasks. This moving metal problem has been solved, or at least reduced to a manageable level by incorporating miniature compensating circuitry at the magnetic sensor [2.2-38].

Electromagnetic emissions from other equipment can also effect the magnetic field and cause error which usually manifests itself as high frequency measurement noise. This type of error can often be eliminated or reduced by properly synchronizing the magnetic system with the offending electro-magnetic source.

It is virtually impossible to accurately predict the performance that will actually be achieved in a particular environment containing the error sources discussed. Empirical testing is required to characterize performance.

2.2.2.5.4. Magnetic Tracking - Prognosis

The problems associated with magnetic tracking still result in significant inconvenience and expense when implementing a magnetic system in a particular environment, especially a military aircraft. Further work is warranted to reduce these problems. Current state of the art does, however, allow the problems to be managed successfully in most cases. Magnetic tracking technology is relatively mature, has been militarized, and offers the best overall head tracking performance available at this time. It is likely to be the predominant head tracking technique for the next generation of military head coupled systems.

2.2.2.6. Safety

General safety issues associated with helmet mounted equipment are discussed under *Some Proposed Applications of Alternative Controls* (section 4.1), and eye safety as relates to optical sources is discussed in the review of *Eye-Based Control* technology (see section 2.3.2.6).

There are no widely accepted standards for exposure to magnetic fields of the type produced by magnetic head trackers, and the biological affects are not well understood. Exposure is usually measured in terms of flux density (field penetration) in the airspace that a person will occupy. The units of measure are gauss (G) in the CGS system and tesla (T) in the SI system ($1\ G = 10^{-4}\ T$).

Flux density from the earth magnetic field varies from about 670 mG at the poles to 330 mG at the equator. Most magnetic head tracking devices do not produce maximum flux density levels that significantly exceed the earth field, although the earth field is static while the field produced by tracking devices is not. It seems very likely that biological effects are dependent on field frequency, but the nature of this dependency is unkown.

Most work on magnetic field safety has focused on 60 Hz power lines, although there are no universally accepted safety standards in this case either. The American Conference of Governmental Industrial Hygienists (ACGIH), for example, recommends an occupational exposure limit of 600 G (60 mT), based on a somewhat controversial computation of the 60 Hz flux density that will induce electric currents not exceeding those normally occurring in the body [2.2-39, 2.2-40, 2.2-41]

2.2.3. APPLICATIONS TO DATE

2.2.3.1. Military Aviation

2.2.3.1.1. Helmet-mounted Sighting Systems

The idea of providing the pilot of a combat aircraft with a helmet orientation sensing system and a simple monocular reticule display, so that he could designate an external target by moving his head to superimpose the reticule over the target, was devised in the early 1960s [2.2-30].

As shown in figure 2.2-5, these two components formed a helmet-mounted sight (HMS) which was integrated into the weapon control system so that helmet orientation signals were sent directly to the seeker head of a lock-before-launch missile, such as the infra-red sensitive AIM-9L "Sidewinder", and the pilot would listen for the change in audible tone that told him when the missile had locked onto the target. He could then pull the trigger and release the missile. This was in contrast to the normal technique which required the pilot to use more extreme manoeuvres to point the aircraft so that the target was brought within the small field of view of the HUD. Essentially the missile "launch success zone" expanded from a cone of about 10° to one of about 60° half-angle, which enabled him to exploit the inherent missile agility and attain earlier weapon release to win the combat.

Slight sophistications brought further benefits. The signals from the helmet sensing system could also be used to point the aircraft radar so that the target range and range rate could be measured and the target g-level computed. Additional symbols in the reticle projector could then be used to tell the pilot whether the dynamically fluid relationship between the two aircraft represented a robust firing opportunity or merely a transitory chance shot. If the pilot looked away the radar would remain locked to the target, and arrow-shaped symbols alongside the projected aiming symbol could be illuminated to cue the direction in which he should move his head to re-acquire visual contact. A similar cueing arrangement could also help one crewmember point out the target to another crewmember.

Early equipment was developed by Honeywell in the form of the Visual Target Acquisition System (VTAS) which used a MOVTAS-type helmet tracker in conjunction with a simple robust reticle projector [2.2-31]. The latter employed an attenuating parabolic visor and a back-illuminated cross-wire to project a small bright aiming symbol into one of the pilot's eyes. This system enabled teams in a number of government laboratories including Aero Medical Research Laboratory (AMRL) in the US and Royal Aircraft Establishment (RAE) in the UK to conduct simulation studies primarily to assess its operational advantages and flight trials to test whether it could be operated satisfactorily by the pilot.

A number of factors were found to have an appreciable effect on the usefulness of a HMS:

- the brightness and sharpness of the reticle image,
- the size and positioning of the optical exit pupil,
- vibration-induced involuntary head motion,
- the difficulty of voluntary head motion at high-g,
- windscreen/canopy optical distortions, and
- the accuracy, update rate and head box size of the helmet tracking system.

As it was found that the pointing error arising from the

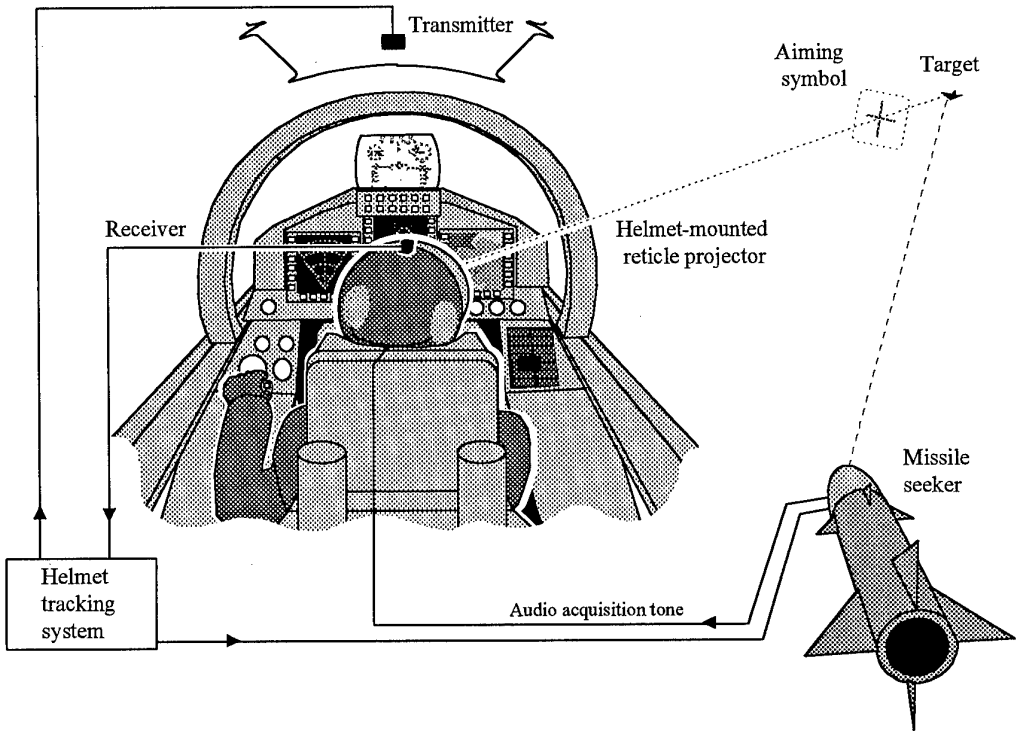


Figure 2.2-5 The basic elements of the helmet-mounted sight (HMS)

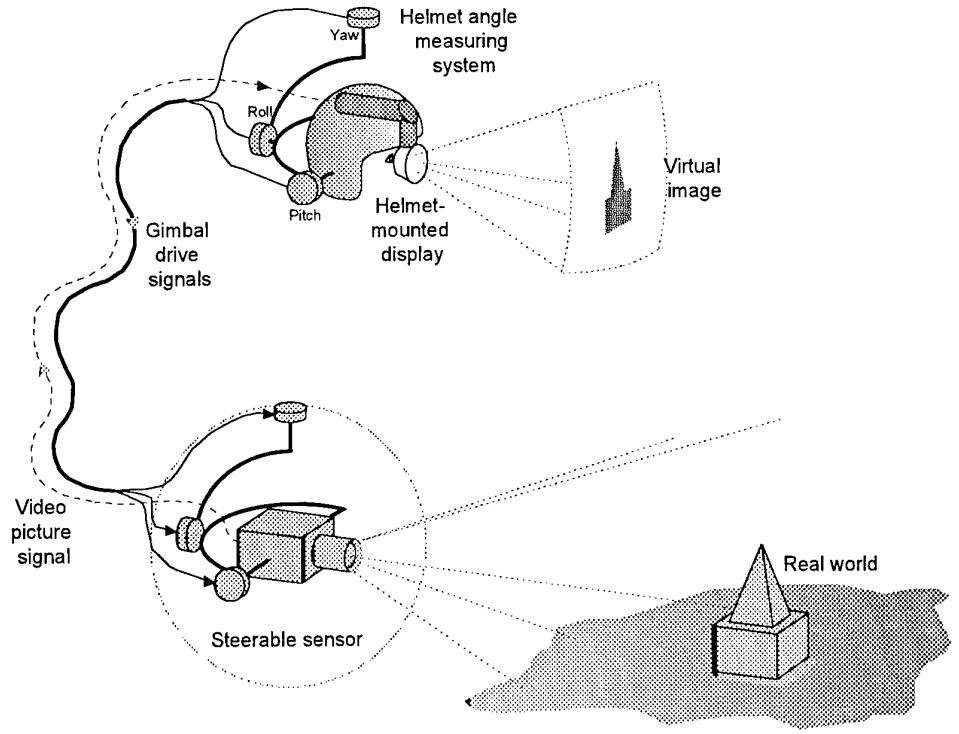


Figure 2.2-6 The idea of a visually-coupled system

combination of technological and human factors was comparable with the small capture field of an infra-red missile, the idea of the HMS was judged to be feasible and worthwhile. This lead to the first service deployment of VTAS in a USAF squadron of F-4 aircraft. Since then HMS systems have been developed by a number of manufacturers and the HMS has become an established facility in combat aircraft operated by the Air Forces of the US, Israel, South Africa and Russia. Systems are also likely to be retro-fitted to other fast jets such as Jaguar, Tornado, F-16 and F-15. The sight will be a standard requirement in all future combat aircraft such as Eurofighter-2000 and Rafale, although in most of these aircraft the aiming symbol will be engineered as one element in a more complex set of imagery.

Note that for this application the helmet position sensing system need only measure the helmet line of sight relative to the airframe, which can be specified by two angles such as azimuth and elevation.

2.2.3.1.2. Visually-coupled Systems

The idea of the visually coupled system (VCS), illustrated schematically in Figure 2.2-6, is a fairly obvious extension of the concept of the HMS to include the feedback of the image from a head-slaved sensor to a helmet-mounted picture-projecting display.

When the field of view of the display matches that of the sensor, the user can have a reasonably normal visual sensation of viewing the world from the sensor location, although the resulting "synthetic vision" is likely to be somewhat limited in scope, quality and sharpness. In general, with suitable communication links and arrangement of the sensor, it is possible to give the user an ego-centric view from an inaccessible, hazardous or remote location, a facility which is currently under investigation for myriad applications ranging from micro-surgery to bomb disposal and tele-robotics. However, it is the use of a sensor, such as a thermal imager working in the atmospheric transmission spectrum between 8 μm and 14μm wavelength, which has been the most notable application. Such a VCS has been developed as the Passive Night Vision System (PNVS) to give the crew of the AH-64 Apache helicopter the means to fly at night and in conditions normally precluded by rain and fog, and not be blinded by missile rocket burn or gun muzzle flash [2.2-42].

The displayed sensor image is invariably overlayed by additional symbols giving flight and weapon aiming information, so the output of the helmet sensing system is simultaneously sent to the symbol generator. It is also available, via the avionics databus, to the rest of the mission and weapon suite to enable the VCS to be used as a HMS. In daylight the system can operate exactly as the HMS described above, using an aiming cross in the centre of the HMD field. At night or in poor visibility, when the sensor image is in use, the pilot can instead move his head so that the image of the target, rather than the directly viewed target, is designated. In the Apache it is also possible for the gunner/co-pilot in the front seat to receive a magnified target image from a narrow field of view sensor so that he can better identify and more accurately designate the target. However, as unwanted head shaking invariably disturbs his aim, because head motion is also magnified, he also has recourse to a head-down display and a joystick to slew the sensor.

The technology employed in the PNVS is a monocular CRT display unit mounted on the side of the helmet, combined with a MOVTAS helmet sensing system. The next generation

of helicopter VCS, such as those integrated into the RAH-66 Comanche and the Franco-German Tiger [2.2-43] will have binocular display systems and electro-magnetic helmet trackers. Similar equipment has been tested satisfactorily in fast jet trials aircraft [2.2-44, 2.2-45], and it is likely to be included in fixed wing combat aircraft which are soon to enter service, such as Eurofighter-2000, and it is under investigation as a means of supplying synthetic vision for future aircraft having windowless cockpits [2.2-46].

Note that for this application the helmet position sensing system must measure the helmet orientation to give correct control over the sensor orientation. Three angular degrees of freedom, such as azimuth, elevation and roll relative to the airframe, are therefore sensed simultaneously.

The requirement to replace an aircraft fixed Head Up Display (HUD) by presenting aircraft stabilized symbology within a helmet mounted display, and maintaining good registrational accuracy with the outside world, calls for head orientation measurement comparable with 1 mradian alignment accuracy of current stationary HUD systems. The HUD application would require this accuracy only over a small forward cone of head pointing angles, and only for the set of HUD applications requiring accurate registration of symbology with real external objects (e.g., delivery of unguided bombs), but head tracking technology needs improvement to achieve this.

2.2.3.1.3. The "Virtual Cockpit"

As summarised in Figure 2.2-7 the idea of the "virtual cockpit" (VC) is to extend the visually-coupled system to its practical limit so that it could provide an integrated and intuitive man-machine interface for all the tasks which make up the pilot's job [2.2-47, 2.2-48]. To enable operations in any external visibility condition, all relevant head-out information for controlling the aircraft, navigating, finding targets, avoiding threats and maintaining tactical awareness would be superimposed directly onto the pilot's normal view of the world or, when this is unavailable, the sensor-derived and computer-generated synthetic substitute for this view. Directional sound cues would provide reinforcement, and the stereoscopic capacity of the binocular display would allow the presentation of cockpit-stabilised 3-D "virtual panels" to convey aircraft systems information and tactical overviews. The idea also postulates that the pilot would control the aircraft flight path and speed using conventional pedals, stick and throttle, and have ready access to HOTAS switches. He would also use a suite of novel controls which are compatible

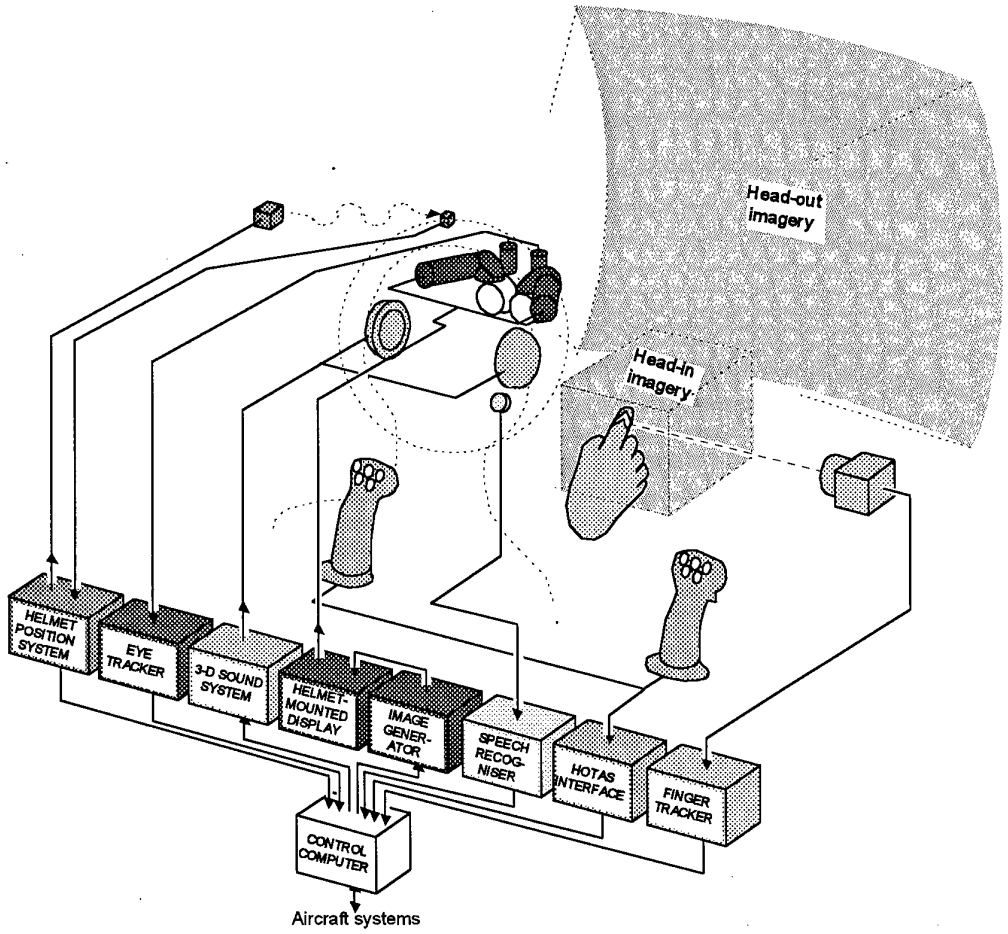


Figure 2.2-7 The likely systems of a virtual cockpit

with virtual imagery.

Note that for this application the helmet position sensing system must measure all six degrees of motion freedom of the head relative to the airframe to enable computation of the precise positions and orientations of the binocular display devices which feed the pair of stereoscopic images into the two eyes. The three degrees of translation (fore/aft, up/down and side-to-side) must therefore be sensed in parallel with the three angular variables. Head position and orientation information is required by the 3-D auditory system and the necessarily powerful image generating computers as well as the head-slaved sensors and the other directable elements of the mission and weapons systems.

2.2.3.2. Non-aviation Uses

The idea of the virtual cockpit is to arrange that as much information as possible occupies its correct position relative to the outside world, where it should be relatively easy to assimilate. For instance the virtual object which shows geographical direction and a symbol which marks a ground radar should not require the extra mental spatial manipulations needed to relate a conventionally-displayed compass bearing or a radar spoke on a head-down warning display to real directions in the the outside world. Thus the ego-centric visual and auditory display of conformal information is aimed at enhancing the pilot's awareness of reality.

This however is the antithesis of the popular idea of immersive "virtual reality" (VR), in which the non-transmissive head-mounted display masks any direct external view in order to obliterate the intrusiveness of the real world and replace it by a computer-manipulated fiction. The technology, which was originally driven by aviation applications, is now being exploited by those interested in commercial and entertainment uses. The requirements of head tracking systems for immersive VR are very similar to those for the VC, without of course the peculiar environmental requirements of aviation, and as the civil market is considerably stronger it is likely that technological development will be driven increasingly by civil needs.

2.2.3.3. Head Based Control of Display Cursors

Most research on head movement based control has either focused on the task of aiming at an external target through a head mounted sight, or on display cursor control.

The use of helmet mounted sights in combat aircraft is beyond the laboratory stage and is now under development for operational use. Studies in this area are cited under *human capabilities and limitations* (section 2.2.1.2), and *military aviation* applications (section 2.2.3.1).

Studies of head control as a computer interface device have often described performance in terms of Fitts' law, a model for human movements. The model predicts movement times proportional to a constant plus an index of difficulty, where the index of difficulty is the logarithm of twice the *distance-to-target/target-width* ratio [2.2-49]. Revised formulations by Welford [2.2-50] and Mackenzie [2.2-51] have modified the index of difficulty by adding a constant to the ratio instead of multiplying by 2.

Jagacinski and Monk [2.2-53], for example, found that time required to aim a helmet mounted sight was adequately described by the Welford formulation [2.2-50] of Fitts law.

Studies of display cursor control using head motion have shown that best performance may be achieved with gains of less than unity; in other words, using relatively large head motions to produce relatively small cursor motions.

Radwin [2.2-53] and Lin et. al. [2.2-54] used an ultrasonic head position measurement device to test head control of a computer screen cursor. The task was to move a cursor from a central home position to a circular target, and hold it on target for at least 62.5 msec. They found head control to be slightly slower than mouse control for the same task. Lin et. al. [2.2-54] found optimal performance of the head position control task with head control gains of 0.3 to 0.6. Fitts law parameters are reasonably consistent with those reported by Jagacinski and Monk.

Another study using a magnetic head tracker to implement a cursor positioning task, produced results that are reasonably consistent with the studies cited above in terms of Fitts law parameters and comparison with mouse control [2.2-55].

Spitz [2.2-56] used an inclinometer type device to control a cursor with head tilt instead of head rotation. He found that target acquisition was slower than that reported by Jagacinski and Monk, but could be adequately described by Fitts' law.

2.2.3.4. Head Tracking in Conjunction with Other Control Techniques

Head tracking is almost always used in conjunction with head mounted eye tracking. Head mounted eye trackers measure gaze direction with respect to the head. When gaze must be must be measured with respect to the cockpit (or other external environment), it is also necessary to measure head position and orientation. This application is discussed further in section 2.3 (*Eye Based Control*).

Head and Gaze tracking have also been used experimentally as the two components of a multimodal cursor positioning control [2.2-55]. Gaze was used as a gross positioning control. For fine positioning, the user was then allowed to use a manual button to switch to head control with less than unity gain. It was concluded that it would probably have been more effective to use manual control as the fine positioning modality. It would also be possible to use head control with unity gain as the gross positioning modality, and a manual technique for fine control. This type of strategy has potential utility in situations where display clutter makes it difficult to simply find the cursor. If a head tracker (or gaze tracker) is used to place the cursor within the immediate field of view for subsequent manual fine positioning, search time is reduced.

Section 2.3.3.3 (under *Eye Based Control)* cites several studies in which gaze measurements are used to provide contextual information as well as positioning information, in conjunction with voice and gesture control modalities. For example, if a user says "zoom", the system knows to zoom the display currently being viewed, about the point being fixated within that display In all of these cases, head control can theoretically be used in place of gaze tracking (combined eye and head tracking). Rather than simply "looking", users would be required to use their head to aim a visor mounted reticule at the desired point of reference. This would be a far less natural task than simply "looking", but head tracking is currently a much more mature technology for operational environments than is eye tracking.

Table 2.2-I Summary of Major Head Tracking Techniques

Method	Major Characteristics	Typical Performance	Status
Mechanical	• Good accuracy • High bandwidth • Low cost • Subject to inertial forces and mechanical damage • Takes up a lot of cockpit space • Mechanical linkage between helmet and cockpit is undesirable (ejection and fast egress problems)	• accuracy: ~5 mm; ~0.2° • update rate: >500 samples/sec • (can vary significantly with specific implementation)	• Has seen operational in-flight use in the past (usually on helicopters for 2 degree of freedom application) • Future use will probably emphasise ground based simulation, R&D, use on helicopters or transports when very low cost system needed.
Inertial	• High bandwidth • Poor static accuracy (requires time integration of accelerations and angular velocities)	• accuracy: ~0.1-1°/sec: ~0.002-0.2 m/sec^2 (not appropriate for static measurement) • update rate: >500 samples/sec	• Potential use in conjunction with other techniques that have good static accuracy.
Acoustic	• Moderate Accuracy • Moderate to poor bandwidth • Echo and blockage problems • Environment noise interference problems • Effected by air temperature and motion	• accuracy: ~5 mm; ~0.5° • update rate: ~30samples/sec	• Requires further work to match optical and magnetic system performance • Systems currently in production are intended primarily for ground based virtual reality applications. • A system is available commercially which combines acoustic and inertial techniques
Optical	• Good accuracy • Moderate to poor bandwidth • Stray IR interference problems (especially from sunlight) • IR emissions may interfere with other cockpit systems that use IR. • Camera mounting problems (multiple cameras must be properly positioned) • Line of sight interference problems	• accuracy: ~1 mm; ~0.2° • update rate: 30 samples/sec	• Mature technology • Military versions available (have seen operational use). • Currently under-perform magnetic systems at similar price
Magnetic	• Very good accuracy • Moderate bandwidth • Large motion box • Metal (including helmet mounted metal) interference and electromagnetic emission problems have largely been solved for most environments, but create expensive and time consuming installation and calibration requirements.	• accuracy: ~1 mm; ~0.1-0.2° • update rate: ~120 samples/sec	• Mature technology • Military versions available • In current operational use • Further accuracy improvement might enable implementation of head mounted HUD

2.2.4. REQUIRED ENHANCEMENTS AND PROGNOSIS

Head tracking devices are a relatively mature technology compared to other enabling technologies for "alternative control" techniques.

Optical devices and both AC and DC type magnetic devices providing full six degree of freedom head position measurement are available in militarised configurations. These devices are in current use, although to a limited degree, in military aircraft.

Improvements are warranted to better handle potential interference conditions (e.g. sunlight for optical systems and moving metal for magnetic systems) and to provide better temporal response. In the case of magnetic systems the interference conditions can often be adequately handled but only with time consuming and expensive calibration procedures. Milliradian accuracy in operational environments would allow an expanded role for head tracking (e.g. head mounted HUD). Magnetic systems are making gains on this benchmark, but it has not yet been reliably achieved.

A summary of head tracking methods is provided in table 2.2-I.

2.2.5. REFERENCES

2.2-1 Glanville, A. D., and Kreezer,G., "The maximum amplitude and velocity of joint movements in normal male human adults", Human Biology, 9, 1937, p 197.

2.2-2 Hertzberg, H. T. E., "Human Anthropology" in VanCott, H. P. and Kinkade, R. G., (Eds) "Human

Engineering Guide to Equipment Design", American Institutes for Research, Wahington D.C., 1972.

2.2-3 Durlach, N. J. and Mavor, A. S., "Virtual Reality Scientific and Technological Challenges", Washington, D.C., National Academy Press, 1995, pp188-204.

2.2-4 Leigh, R. J., and Zee, D. S., "The Neurology of Eye Movements", Philadelphia, F. A Davis Company, 1983, pp 109-123.

2.2-5 Bizzi, E., "Eye-head coordination". In Brooks, V. B. (Ed) "Handbook of Physiology, The Nervous System", Sect 1, Vol 2, Part 2, Ch29, Bethesda, MD, American Physiological Society,. 1981, pp 1321-1336.

2.2-6 Bizzi, E., Kalil, R. E., and Morasso, P., "Two modes of active eye-head coordination in monkeys", Brain Research, 40, 1972, pp 45-48.

2.2-7 Mourant. R. R. and Grimson, C. G., "Predictive head-movements during automobile mirror sampling", Perceptual and Motor Skills, 44,. 1977, pp 283-286.

2.2-8 Barnes, G. R. and Sommerville, G. P., "Visual target aquisition and tracking performance using a helmet-mounted sight", Aviation, Space, and Environmental Medicine, April, 1978, pp 565-572.

2.2-9 Wells, M. J. and Griffin, M. J., "A review and investigation of aiming and tracking performance with head-mounted sights", IEEE Trans on Systems, Man and Cybernetics, SMC-17, 2, 1987, pp 210-221.

2.2-10 Sandor P. B., Leger A., "Tracking with a restricted field of view: performance and eye-head coordination aspects", Aviat. Space Environ. Med; 62, 11, 1991, pp 1026-31.

2.2-11 Viviani P., Berthoz A., "Dynamics of the head-neck system in response to small perturbations: Analysis and modeling in the frequency domain", Biol. Cybernetics 19, 1975, pp 19-37.

2.2-12 Griffin, M. J., "Vertical vibration of seated subjects: Effects of posture, vibration level and frequency", Aviation, Space and Environmental Medicine, 46, 1975, pp 269-276.

2.2-13 Rowlands, G. F., "The transmission of vertical vibration to the head and shoulders of seated men", Royal Aircraft Establishment Technical Report TR-77068. Farnborough, England, 1977.

2.2-14 Lewis C. H., Griffin M. J., "Predicting the effect of vibration frequency and axis and seating conditions on the reading of numeric displays", Ergonomics, 23, 1980, pp 485-507.

2.2-15 Furness T. A., "The effect of whole body vibration on the perception of the helmet-mounted display", Ph. D. dissertation, Univ. Southampton (unpublished), 1981.

2.2-16 Tatham, N. O., "The effects of turbulence on helmet-mounted sight accuracies", AGARD CPP 267, 1979.

2.2-17 Fong, K. L., "Maximizing +Gz Tolerance in Pilots of High Performance Combat Aircraft", US Air Force Report AL-SR-1993-0001, December 1992.

2.2-18 Leger A., Sandor P., Clere J. M., Ossard G., "Mobilité de la tête et facteur de charge: approche expérimentale en centrifugeuse", AGARD-CP 471, AMP Symposium on "Neck injury in advanced military aircraft environments", Munich, Germany, 1989.

2.2-19 Leger A., Sandor P. "Désignation de cible sous facteur de charge: intérêt et limites du viseur de casque", AGARD-CP 478, AMP symposium on "Situational awareness in aerospace operations", Copenhagen, Denmark, .11, 2-5 October, 1989, pp 1-10.

2.2-20 Leger A., Sandor P., Troseille, X., "Désignation d'objectifs sous facteur de charge: poursuite de cibles mobile", R.E. N° 32 CEV/SE/LAMAS, 1990.

2.2-21 Leger A., Sandor P., Bourse C., Alain A., "Réponse biomécanique de la tête aux accélérations +Gz: Intérêt pour les études en simulation de combat", AGARD CP-517, "Helmet Mounted Displays and Night Vision Goggles", Pensacola, Fl, 6,-1991, pp 1-9.

2.2-22 Leger A., unpublished observations, 1993.

2.2-23 Kocian, D. F., and Task, H. L., "Visually Coupled Systems Hardware and the Human Interface" In Barfield, W., and. Furness, T. A, (Eds) "Virtual Environments and Advanced Interface Design", New York, Oxford University Press, 1995.

2.2-24 Jarrett, D. N., "Helmet position sensor and loading mechanism", DRA working paper DRA-FS-93-WP892, 1993.

2.2-25 "Operator , Organizational, Direct Support and General Support Maintenance Manual", US Army Technical Manual TM 9-1270-212-14&P, July, 1981.

2.2-26 Foxlin, E.; and Durlach, N., "An inertial head-orientation tracker<with automatic drift compensation for use with HMD's", in Singh, G., Feiner, S., K., and Thalmann, D. (Eds.) "Virtual Reality Software and Technology. Proceedings of the VRST '94 Conferrence", Singapore, 1994, pp 159-73.

2.2-27 Emura, A. and Tachi, S., "Compensation of time lag between actual and virtual spaces by multi-sensor integration", in "Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems", Las Vegas, NV, 1994.

2.2-28 Sutherland, I. E., "A head mounted 3 dimensional display", in "1968 Fall Joint Computer Conference, AFIPS Conference Proceedings", 33, 1968, pp 757-764.

2.2-29 Applewhite, H. L., "A new ultrasonic positioning principle yielding pseudo-absolute location", in Singh, G., Feiner, S., K., and Thalmann, D. (Eds.) "Virtual Reality Software and Technology.

Proceedings of the VRST '94 Conference", Singapore, 1994, pp. 175-83.

2.2-30 Jacobs R. S., Triggs T. J., and Aldrich J. W., "Helmet-mounted display/sight system study" US Air Force Technical Report AFFDL-TR-70-83 Vol 1, 1970.

2.2-31 "An introduction to Honeywell helmet-mounted displays", Avionics Division, Honeywell, 1977.

2.2-32 Kuipers, J. B., "SPASYN -- an electromagnetic relative position and orientation tracking system", IEEE Transactions on Instrumentation and Measurement, IM-29, 4, 1980, pp 462-466.

2.2-33 Raab, F. H., Blood, E. B., Steiner, T. O., and Jones, H. R., "Magnetic Position and Orientation Tracking System", IEEE transactions on Aerospace and Electronic Systems, AES-15, 5, 1979.

2.2-34 Blood, E., "Device for Quantitatively Measuring the relative position and orientation of two bodies in the presence of metals utilizing DC magnetic fields", US patent no. 4,849692, 1989.

2.2-35 Blood, E., "Device for Quantitatively Measuring the relative position and orientation of two bodies in the presence of metals utilizing DC magnetic fields", US patent no. 4,945,305, 1990.

2.2-36 Adelstein, B. D., Johnston, E. R., and Ellis, S. R., "Dynamic Response of Electromagnetic spatial displacement trackers", Presence, 5, 3, 1996, pp 302-318.

2.2-37 Hericks, J., Parise, M., and Wier, J., "Breaking down the barriers of cockpit metal in magnetic head tracking", in "Proceedings of the SPIE Head Mounted Displays", Orlando, FL, April 8-10, 1996.

2.2-38 Brindle, J. H., "Advanced helmet tracking technology developments for naval aviation", in "SAFE association, 33d Annual Symposium", Reno, NV, Oct. 23-25, Proceedings (A96-1671603-54), 1995, pp 34-53.

2.2-39 Lee, J. M.; Chartier, V. L.; Hartmann, D. P.; Lee, G. E.; Pierce, K. S., "Electrical and Biological Effects of Transmission Lines: A Review", U.S. Department of Energy Report DOE/BPA-945, Jun 1989.

2.2-40 Walsh, M. L., and Donnally, K. E., "Power frequency electric and magnetic field exposure and human health", in "35th Cement Indus. Tech. Conf", Toronto, Canada, May 23-27, IEEE Cat. No. 93CH3268-0, 1993, pp 279-88.

2.2-41 "1998 TLVs and BEIs", American Conference of Governmental Industrial Hygenists, Cincinnati, OH, 1998, (ISBN 1-88-2417-23-2), p 142.

2.2-42 Brown T. C., "AH-64 Apache night vision system" in "Night vision '92" Conference, London, 1992.

2.2-43 "Glass cockpit operational effectiveness", AGARD AR-349, 1996.

2.2-44 Church T. O. and Bennett W. S., "System automation and pilot-vehicle-interface for unconstrained low-altitude night attack" in "Combat automation for airborne weapon systems: Man-machine interface trends and technologies", Edinburgh, UK. AGARD-CP-520, 1992.

2.2-45 Lydick L. N., "Head-steered sensor flight test results and implications" "Combat automation for airborne weapon systems: Man-machine interface trends and technologies", Edinburgh, UK. AGARD-CP-520, 1992.

2.2-46 Rolwes M. S., "Design and flight testing of an electronic visibility system" in "Helmet-mounted displays II", SPIE 1290, 1990, pp 108-119.

2.2-47 Furness T. A. and Kocian D. F., "Putting humans in virtual space" The Society for Computer Simulation, Simulation Series, 16, 2, San Diego, CA, 1986, pp 214-230.

2.2-48 Kaye M. G., Ineson J., Jarrett D. N. and Wickham G., "Evaluation of virtual cockpit concepts during simulated missions", in "Helmet-mounted displays II" SPIE 1290, 1990, pp 236-245.

2.2-49 Fitts, P. M., "The information capacity of the human motor system in controlling the amplitude of movement", Journal of Experimental Psychology, 47, 1954, pp 381-391.

2.2-50 Welford, A. T., "The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress", Ergonomics, 3, 1960, pp 189-230.

2.2-51 MacKenzie, I. S., "Fitts' law as a research and design tool in human-computer interaction", Human Computer Interaction, 7, 1992, pp 91-139.

2.2-52 Jagacinski, R .J., and Monk, D. L., "Fitts' law in two dimensions with hand and head movements", Journal of Motor behavior, 17, 1985, pp 77-95.

2.2-53 Radwin, R. G., "A method for evaluating head-controlled computer input devices using Fitts' law", Human Factors, 32, 4, 1990, pp 423-438.

2.2-54 Lin, M. L., Radwin, R .G., and Vanderheiden, G .C., "Gain effects on performance using a head-controlled computer input device", Ergonomics, 35, 2, 1992, pp 159-175.

2.2-55 Borah, J., "Investigation of Eye and Head Controlled Cursor Positioning Techniques", US Air Force report AL/CF-SR-1995-0018, September 1995.

2.2-56 Spitz, G., "Target acquisition performance using a head mounted cursor control device and a stylus with digitizing tablet", in "Proceedings of the Human Factors Society 34th Annual Meeting", 1990, pp 405-409.

2.3. EYE-BASED CONTROL

2.3.1. INTENTION OF THE TECHNOLOGY

The principal objective is to measure the eye line of sight so that air crew can designate targets or features in the external world, and interact with objects or switches presented in virtual display systems. Line of sight may also be used as context information by other control modalities such as voice or gesture recognition.

The same type of measurements may also be useful for pilot state monitoring, but those applications are beyond the scope of this report.

2.3.1.1. Relevance

Use of the eye for designation is intended to exploit the naturalness, speed and accuracy of visual fixation and tracking. In comparison with head pointing it offers several benefits; the user may be able to cover a wider angular envelope, and the combination of speed, accuracy and scope is less likely to deteriorate under turbulence-induced vibration, or during high-g combat maneuvering.

Where the line of sight is measured relative to the headgear, it is also necessary to measure the headgear position and orientation to compute the eye line of sight with respect to the airframe. Measurement of one eye is sufficient, but it may be necessary to provide the crew member with a convenient (HOTAS) switch or voice command to mark instants of designation.

2.3.1.2. Human Capabilities and Limitations

Each eye is rotated in its socket by a set of antagonistic muscles to bring high resolution foveal vision to bear upon features in visual space. The lines of sight and erectness are also controlled reflexively by the vestibular system to give stability against head motion and by binocular image differences so that both eyes converge on the same feature.

When examining a stationary scene, both lines of sight are simultaneously held steady for short periods (usually 200 - 600 msec), called fixations, to bring a feature of interest within the approximately 1° angular range of the fovea. Miniature eye movements of up to several minutes of arc do occur during the periods of "fixation", but are not perceived. A more detailed description of miniature eye movements, called flicks, drift, and micro-saccades, can be found in references [2.3-1] and [2.3-2]. A very thorough review of eye movements in general can be found in Hallet [2.3-3].

Rapid jumps, called saccades, move the eye between fixations. Saccades usually reach velocities of 400-600 °/sec, and last 30 - 120 msec. Vision is significantly suppressed during this period. Although saccades can be as large as 50°, they are more commonly 1 - 20°. If a target appears in peripheral vision, it takes a minimum of about 100 msec for a saccade towards the target to be initiated.

When observing a slowly moving object the lines of sight usually track smoothly, but this pursuit reverts to fixations and saccades when the object moves faster than about 30°/sec. Without specific training, smooth eye movements are only possible when following a smoothly moving target or compensating for head movement.

Visual acuity is best on the foveal region of the retina, and within the fovea is best near the very center. People therefore direct the visual axes of the eyes (axes passing through the center of each eye lens and fovea) to objects that they want to see clearly; however, there may be a foveal "dead zone" or "indifference threshold" on the order of about 0.3 degrees visual angle for fixation of stationary targets [2.3-4]. Attention can be shifted within the foveal region, and even outside of the foveal region if the target of interest falls within acuity limits of peripheral vision [2.3-5, 2.3-6, 2.3-7]. Furthermore, foveation accuracy falls off markedly if a person attempts to maintain fixation for several seconds [2.3-2, 2.3-8], when tracking a moving target [2.3-3, 2.3-9, 2.3-10, 2.3-11], or during rapid head movements [2.3-12]. Thus, even if we could measure direction of the visual axis with infinite accuracy we would not always have perfect knowledge of point of regard. If a person consciously attempts to fixate a small, stationary, target, for a short time, while holding their head steady, we can probably assume the visual axis to be within 0.3 degrees of the target.

The eyes can move over an angular range of about ±50° horizontally, and about +40°, -60° vertically with respect to the head. Normally, eye movement with respect to the head remains within about ±15°-20° [2.3-13, 2.3-14]. Gaze shifts beyond the central 20° field are usually, although not always, accompanied by head rotation. Horizontal eye rotation with respect to the head of more than about 40° from the central position becomes quite uncomfortable if maintained for several seconds.

When using eye movements to enable control or interaction techniques there are several general characteristics of human eye movement and gaze behavior that should be kept in mind:

- The normal fixation/saccade pattern of visual scanning can be thought of as a continual series of snap shots that are used to create a mental image of the visual environment; however this is usually an unconscious process. Perception of the environment is of the "single picture" formed in the brain.
- People are not accustomed to controlling things with their gaze, and are accustomed to being able to glance at things without causing some action to occur.
- It is difficult and annoying, although possible, to maintain steady fixation on a single target for much more than a second. Fixations of several hundred milliseconds are most natural. There is also a strong tendency to make quick glances at other nearby targets during unnaturally long fixations.
- The eye is drawn to features, and it is very difficult to fixate a blank spot.
- Feedback of gaze position (presentation to a person of their own gaze point as measured by an eye tracker, and often referred to as secondary visual feedback) must be handled carefully. Continuous feedback of gaze position, if not perfectly accurate and up to date, can sometimes be distracting instead of helpful. If the displayed indicator is slightly displaced from the central line of gaze there may be a tendency to continually try to

look at it, leading to a positive feedback loop. Techniques for dealing with this are discussed further in section 2.3.3.4.

2.3.2. OVERVIEW OF APPROACHES

Devices to measure eye line of sight were developed primarily as laboratory tools for ergonomic and psychological research. Only very gradually, over the past 10 years, have these devices begun to find usage as applied tools. The current generation of devices has probably not yet reached the level of true practicality for applied use in aerospace cockpit environments, but this does appear to be a reachable horizon. Several terms are very frequently used in relation to devices that measure eye line of sight or related quantities. *Line of gaze* is the imaginary straight line extending from the center of the fovea, through the center of the eye lens and out to infinity. In other words, it describes the location in space of the visual axis. *Point of gaze* refers to the point whose image actually forms at the center of the fovea. It is the intersection point of the line of gaze with a visible surface. *Point of regard* implies the point in the visual environment that the subject is actually paying attention to. Usually this is the same as point of gaze, but not always. *Eye tracker* refers to a device that measures eye orientation (pointing direction) with respect to some measurement reference frame. Often this reference frame is the subject's head gear, but sometimes it may be an optics package, or Helmholtz coil mounted to some other surface in the environment. In some cases an eye tracker alone is sufficient for measuring line of gaze or point of gaze. For example, an optical pupil to corneal reflection method eye tracker, with an optics package mounted to the bottom of a display, may be quite sufficient to measure point of gaze on the display surface. In other cases additional input from a head tracker or a navigation system is required to determine line of gaze or point of gaze in the relevant environment. The entire system for determining line of gaze or point of gaze can be referred to as a *gaze tracker.*

Although the *point of gaze* can be described unambiguously by co-ordinates on the intersected surface, the direction of the *line of gaze* is invariably, but somewhat loosely, described as a "horizontal" and "vertical" pair of angles. As it is usually necessary to combine eye direction with head and aircraft orientation, using the relationships between eye, head, cockpit, aircraft and earth reference frames summarized in Appendix E, it is essential to have a rigorous understanding of what "horizontal" and "vertical" signify. Here it should be noted that alternative co-ordinate systems, named after Helmholtz, Fick and Listing are used to describe the orientation of the eyeball with respect to the head by experimenters interested in different aspects of eye motion [2.3-15]. Of these, Listing's is equivalent to the familiar set of azimuth, elevation and bank angles used for aircraft orientation, and, since it is unnecessary to measure "torsion" of the eye (analogous to aircraft bank angle) to specify the line of gaze, in the absence of explicit indications it is usually assumed that "horizontal/vertical" signify "azimuth/elevation".

Eye tracker performance is often described in terms of the following parameters. *Accuracy* is the expected difference between measured eye line of gaze and true eye line of gaze, usually expressed in terms of visual angle. *Precision* (repeatability) is the expected difference in repeated measurements of the same true eye line of gaze. *Linearity* is the degree to which a change in the measurement is proportional to the actual change in eye angle, and is usually expressed as a percent of the eye angle change being measured. Stated another way, linearity is the amount that a plot of measured values versus actual values is expected to deviate from a straight line. *Resolution* is the smallest change in eye angle that can be reported by the device. *Range* is the amount of eye motion that can be measured, usually specified in degrees visual angle. Range may be specified with respect to the head gear or with respect to the external environment (e.g. airframe), depending upon the device reference frame. *Update rate* is the frequency with which data samples are measured and reported, usually as "samples/second". *Transport delay* is the amount of time that it takes data to travel through the system and become available for use. *Latency* (or *throughput* as defined by Kocian and Task [2.3-16]) usually refers to the amount of time required to accurately reflect a change in the quantity being measured. It is influenced by pure transport delay and also by dynamic operators (for example, a low pass filter) in the signal path. *Bandwidth* is the range of sinusoidal input frequencies that can be processed by the system without significant attenuation or distortion.

The predominant eye tracking techniques can be classified as *electro-oculographic*, *scleral coil*, and *optical* methods. The optical classification is the largest of these categories, and contains multiple sub-categories. The following subsections discuss each major type of eye tracking technique followed by a discussion of calibration and safety issues.

2.3.2.1. Electro-Oculography

2.3.2.1.1. Physiological Mechanism

The idea of placing electrodes close to the eyes to detect changes in the eye pointing direction arose originally because movement of the eyes, along with activation of any head or jaw muscles, is a considerable source of noise in electro-encephalography.

The phenomenon is thought to arise because the retina at the back of the eye develops a small negative electrical charge relative to the front surface of the cornea, probably as a result of its higher metabolism [2.3-1]. This corneo-retinal potential, or electrostatic dipole, depends mainly upon the ambient light level, and its maximum varies between individuals from about 0.4 to 1.0 mV. The effective negative pole is close to the optic disk, about 15 degrees from the macula, so the electrical dipole is not aligned precisely with the eye's optic axis.

2.3.2.1.2. Description of Technique

When a pair of electrodes is placed on the surface of the skin on either side of the eye, the corneo-retinal dipole induces zero differential voltage when the dipole axis is about midway between electrodes. A change of about 20 μV/° results when the eye is rotated towards one of the electrodes. Vertical and horizontal movement are sensed respectively by one pair of electrodes attached above and below the eye and another pair attached on either side close to the inner and outer canthi. Independent measure of the left and right eyes is therefore possible in principle but, as shown in Figure 2.3-1, to obtain a general indication of eye-pointing direction it is more common to mount a single pair at the outer canthi of both eyes to sense their combined horizontal effect, and sense

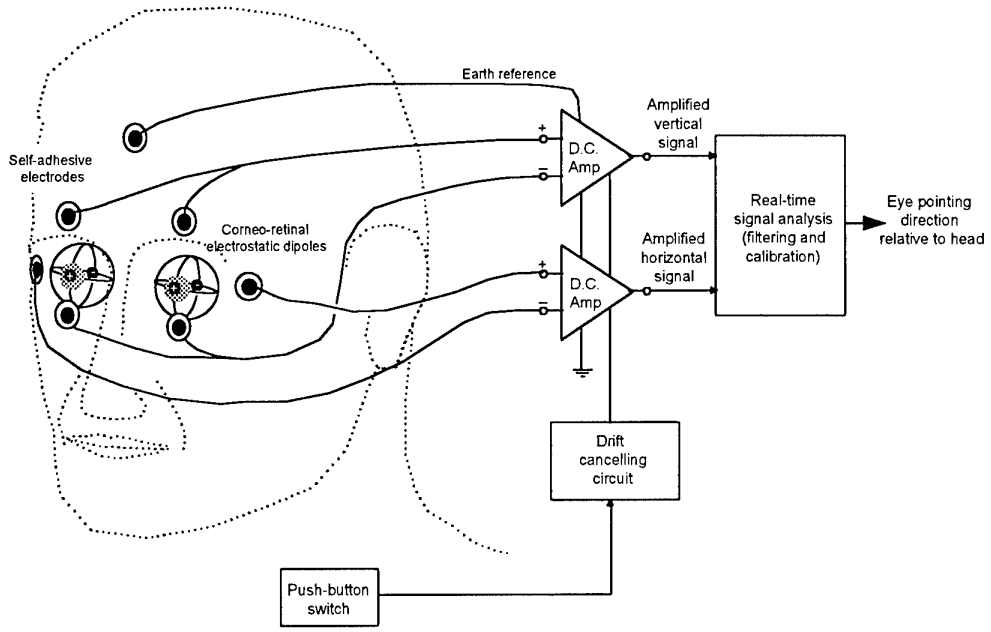


Figure 2.3-1. Schematic illustrating the electro-oculographic (EOG) technique for measuring eye motion

either the vertical motion of just one eye or connect the two vertical sets in parallel [2.3-17].

Small commercial (silver)+(silver chloride) skin electrodes, normally used for monitoring the heart functioning in babies, are commonly employed to minimize electro-chemical artifacts. The skin is cleaned and de-greased with an alcohol swab. Then the contact surface is wetted with conductive saline gel and the electrode is fixed in place using the adhesive backing ring. The short leads are connected to high gain, high (>1 MΩ) impedance, low noise, low drift differential amplifiers having a bandwidth from zero to about 100 Hz. Calibration is required to scale and map the EOG signals to coordinates of gaze with respect to the head.

Although the magnitudes and polarity of the EOG signals are evidently related to the directions of regard relative to the head, the uncontrollable diurnal variations are compounded by non-linearity's, cross-coupling, noise and drift. Drift is minimized by connecting a "reference" electrode, sometimes sited at the center of the forehead, to the amplifier ground. In any event, drift must be countered to keep the signals within the dynamic range of the amplifiers, and a convenient facility is usually provided to enable the subject to re-zero the output when periodically looking straight ahead. Noise is minimized by filtering and by avoiding unnecessary muscle activity, particularly clenching of the jaw. Other artifacts such as short spikes which accompany vertical movements, thought to be due to the change in surface conductivity or myographic potential from eyelid movement, are more difficult to remove. Non-linearity's and cross-coupling are inherent and, depending upon the range and accuracy sought, can only be treated by a painstaking calibration. Note, however, that the calibration must be completed quickly as it is also disturbed significantly by the inherent variations and drift.

2.3.2.1.3. Performance

This analog technique offers a potential combination of fine sensitivity, short delay and unlimited range. However, the low-pass filtering needed to reduce noise and transient artifacts compromises both sensitivity and speed. Accuracy depends principally on the dipole strength, and its uncontrollable variation over time, including the drift incurred by the combination of electrodes and amplifier. Most of these factors vary between individuals, and subjects range from those for whom the EOG signal is weak, variable and unreliable to those for whom it is strong and relatively consistent.

Extrapolating from laboratory measurements by Shackel [2.3-17] and in flight tests conducted in a Jaguar aircraft [2.3-18] it seems reasonable to conclude that EOG measures in a cockpit environment might allow inference of eye pointing direction relative to the head with an expected error between about 3° and 7°, assuming some form of filtering and frequent re-zeroing.

2.3.2.1.4. Practical Problems

There seem to be few practical difficulties in attaching unintrusive electrodes to a pilot and obtaining satisfactory EOG signals. The technique accommodates the wearing of a helmet and oxygen mask, and it should be compatible with other devices such as night vision goggles and any future integrated helmet systems. It has the benefit of generating measurements with closed eyes.

The principal operational problems are that the technique requires frequent re-zeroing and calibration, and the attainable accuracy is likely to vary considerably between individuals. Unfortunately, unlike scientists conducting

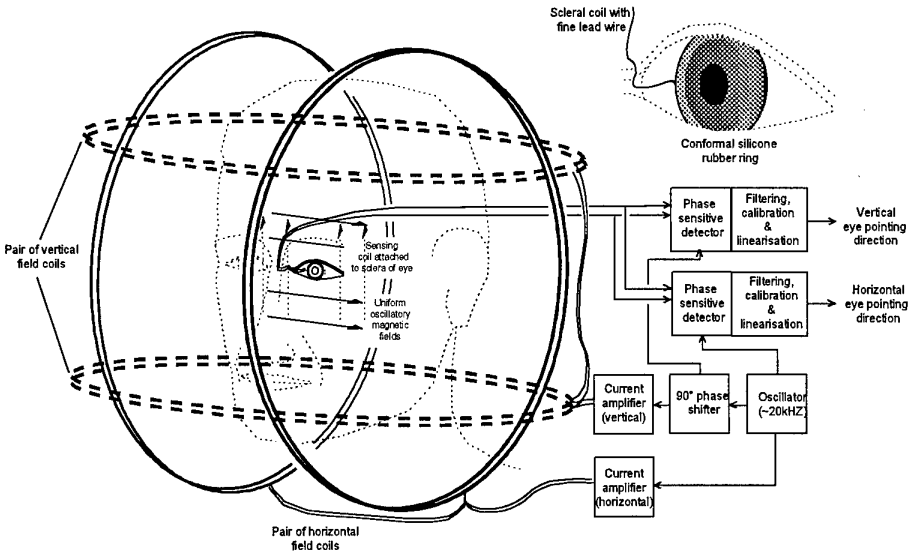


Figure 2.3-2. Schematic illustrating the scleral search coil method of measuring eye movement

laboratory research, aircraft operators cannot select crew having a strong corneo-retinal potential.

2.3.2.1.5. Prognosis

The demonstrated practicality in a combat aircraft and the potentially unrestricted range and robustness of the technique make it eminently suitable for general aircraft use. However even for "good" individuals the accuracy will remain inadequate for target designation.

It is most likely that EOG measurement would be included in a suite of alternative control mechanisms to complement other techniques. For instance, it could supply measurements beyond the range of an optical eye tracker, which in turn could be arranged to keep it in better calibration. EOG measurements could also help to remove eye motion artifacts from EEG signals and supply data about eye blinks and motion-induced nystagmus which would help infer the pilot's state.

2.3.2.2. Scleral Coil

The scleral coil is one of a class of techniques which require attachment of a sensing element to the subjects eye. Early work by Orchansky [2.3-19] and Delabarre [2.3-20] showed that it was possible to fix a metal ring or a glass shell to the suitably anaesthetized cornea. It was the later experiments of Marx and Trendelenburg [2.3-21] who studied the motion of a tiny mirror on such a shell, by reflecting a beam of light from a nearby source onto a moving photographic film, showing that the eye constantly jittered by about 4 or 5 minutes of arc during fixation. Other investigators, invariably using a contact lens as the mount, have attached other sensors such as mechanical levers, miniature lamp, birefringent material or strain gauges to the eye.

2.3.2.2.1. Description of Technique

The use of a magnetic search coil was developed by Robinson [2.3-22]. As illustrated in Figure 2.3-2, to measure horizontal eye motion, the subject sits with his head inside a pair of co-axial Helmholtz coils, which set up a horizontal uniform oscillatory magnetic field, and this induces a voltage in a very fine induction coil which is attached to the eye. The sensor coil is embedded in a shallow ring of silicone rubber, the inner surface of which is slightly hollow, so that it adheres to the limbus by capillary action and suction and remains concentric with the corneal bulge. The strength of the induced signal varies with the sine of the horizontal angle between the scleral coil axis and the main field, and phase sensitive detection is used to find a signal which is exactly in phase with the excitation. A second set of field coils, perpendicular to the first set but energized with a signal at 90° phase shift is used to excite a response to vertical rotation which is also detected synchronously.

The technique is reasonably immune to asynchronous magnetic field distortions from external causes, and the synchronous defects from nearby conductive or ferromagnetic material can be removed by sensing the distortions using a secondary detector which is stationary relative to the main field coils. Complete scleral coil systems are available commercially [2.3-23]. Such systems are available with several sizes of field coil, usually wound on cubic formers, and are capable of sensing eye torsion, rotation about the line of sight, using a second orthogonal sensing coil inside the scleral ring. Also, the pointing direction of both eyes can be measured with respect to the fixed excitation coils.

2.3.2.2.2. Performance

Following a simple calibration to define the initial reference orientation of the eye, the rotations can be measured to a resolution of about 1 arcmin over a range of about ±15° to an accuracy of about 1% of the range. Speed is limited only by the excitation frequency and any filtering; a bandwidth of 0 to 200 Hz would be reasonable. In the laboratory the technique is dependable and accurate, and it is intrinsically impervious to subject differences, light level variation and the position of the eyelid.

2.3.2.2.3. Practical Problems

Installation in an aircraft presents two main problems. Firstly, as it would probably be impractical to engineer excitation coils which are large enough to allow normal head movement, it is more attractive to consider mounting such coils on the pilot's helmet and measure eye direction relative to the helmet. Although coils could be made to be tolerably lightweight, siting them to produce an even field strength at the eye, under the perturbations caused by the other electro-optical components on the helmet, would remain problematical. No work to pursue this approach is known.

Secondly, the wearing of a coil embedded in a scleral ring or a contact lens is slightly, but distinctly invasive. Most individuals would require a drop of eye anesthetic when the rings are offered up to the eye. The thin wires connecting the coil to the sensing electronics, although brought out across the nasal corner of the eye and taped securely to the side of the nose, would be apparent with each blink. It is perhaps the concern about such an intrusion, and the possibility of the coil or wires becoming dislodged, or a corneal oedema which seems to make the scleral coil unacceptable for operational use. On the other hand, since disposable contact lenses may become the favored means for correcting refractive errors in aircrew [2.3-24], the attachment of additional fine structures to the eye may also become acceptable, and this concern may be unwarranted.

2.3.2.2.4. Prognosis

Although the scleral coil is an excellent laboratory tool, the anticipated problems of integrating it into a cockpit remain. Unless other approaches prove inadequate, it is unlikely to receive the necessary testing and development.

2.3.2.3. Optical Techniques

2.3.2.3.1. General Description

Optical eye tracking techniques make use of optically detectable eye features and geometry to determine the orientation of the eye ball.

The following features, illustrated in Figure 2.3-3 are most commonly used :

- *Limbus* -- the boundary between the colored iris and white sclera.
- Pupil -- the opening in the iris (aperture of the eye)
- Corneal reflection (CR), or first Purkinje image -- mirror reflection of an external source from the outer surface of the cornea
- 4th Purkinje image (4PI) – mirror reflection of an external source from the rear surface of the eye lens.

Eye ball orientation can be computed from the position of a single feature if the sensor is assumed to be rigidly fixed to the head or if sensor position with respect to the head can be independently measured. Single feature eye trackers usually rely on the limbus, the pupil, or a corneal reflection. Sometimes the lower eye lid is also used as a low fidelity indicator of vertical eye position.

Position of a single feature alone will not distinguish rotation of the eye ball from movement of the sensor. As shown by Figure 2.3-3, multiple landmarks located at different radii from the center of the eye ball will appear to move with respect to one another as the eye rotates, but will move together when the sensor translates. By differentiating between eye rotation and translation with respect to a sensor, dual feature techniques minimize errors due to shifting of head mounted optics, and also allow use of non head mounted optics. It is also important to note that the response of the dual feature technique is described by a sine function which is maximally sensitive at small angles. Dual feature systems usually use the pupil and corneal reflection, or the corneal reflection (CR) and 4th Purkinje image (4PI). The pupil forms a landmark near the eye ball surface (about 9.8 mm from the eye ball center), the CR behaves as would a land mark at the same radius from eye ball center as the corneal

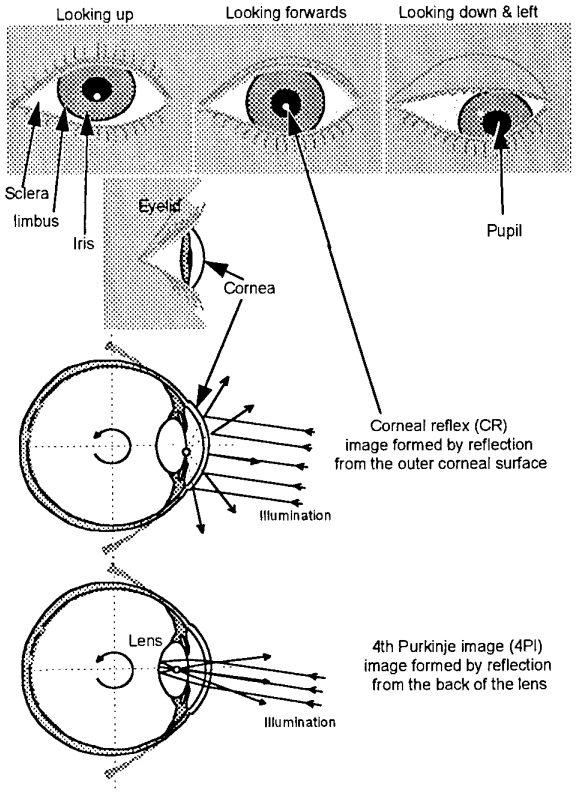


Figure 2.3-3 Eye features exploited by optical eye tracking systems

center of curvature (about 5.6 mm), and the 4PI appears to move the same amount as the posterior lens surface center of curvature (about 11.5 mm from eye ball center).

In theory, feature shape can also be used to distinguish rotation from translation since rotation will cause apparent change in the feature dimension that is perpendicular to the rotation axis. When viewed from an angle, for example, the round pupil appears to be an ellipse. The relation of the major and minor axes of the pupil ellipse is the cosine of the angle between the sensor and the optical axis of the eye. Because the cosine function has a very shallow slope at small angles, the sensitivity of this technique tends to be poor. Symmetry characteristics also create ambiguity about the direction of rotation.

Optical eye trackers usually include an illumination source to help produce a suitable image for the optical sensor. This source is most often restricted to the near infra red spectrum to avoid disturbing the user while still providing effective illumination for the sensor. Solid state optical sensors usually have good sensitivity in the near infra red (IR) although it is just beyond the visible spectrum. Care must be taken to keep the illumination at safe levels, and this is addressed in more detail further on. Head mounted illumination sources are usually solid state emitters (LEDs or laser diodes). Illumination sources that are not head mounted may be high power solid state emitters, arrays of solid state emitters, or filtered incandescent sources.

Sensors used by optical eye trackers include the following:

- *solid state quadrant or bicell detectors* – provide analog information that is monotonically related to very small displacements of a spot of light from the center of the detector in one (bicell) or two (quadrant) dimensions. Displacements must be smaller than the diameter of the light spot.
- *lateral effect photo diodes (position sensitive detectors)*– provide analog information proportional to the one or two dimensional location of the incident light center of gravity.
- *pairs or very small arrays of individual solid state photo detectors* – provide a small number of analog light intensity signals.
- *large linear arrays of solid state photo detectors* – provide one dimensional gray scale data
- *Two dimensional solid state arrays (optical RAM, CCD, and CID)* – provide two dimensional binary (optical RAM) or gray scale (CCD and CID) images.

A generic optical gaze tracker, as shown in figure 2.3-4, consists of an optical unit including an illumination source and sensor with related optics, electronics to operate the sensor, and a processing unit to compute gaze with respect to the optics. If the eye tracker optics are head mounted, it must be augmented with a head tracker in order to implement a gaze tracking system capable of measuring point of gaze in the cockpit or workstation (see discussion of reference frame relationships in Appendix E). A head mounted eye tracker *need not* be augmented with a head tracker in order to measure point of gaze on a head mounted display.

As previously discussed, dual feature eye tracking techniques permit the use of non head mounted optics which directly measure line of gaze with respect to the cockpit or work station, without requirement of a head tracker. In this case, however, the system must successfully find the eye amid the clutter of the environment as the user moves about, and in the presence of any head mounted optical components such as head mounted displays.

Quadrant detectors, lateral effect photo diodes, and small arrays of photo-detectors provide information with low spatial bandwidth content. This information can usually be processed very quickly to achieve high temporal bandwidth with minimal requirement for digital processing. Large linear and two dimensional arrays provide information that is very rich in spatial content (e.g., a real gray scale image), but systems that use these sensors typically require more digital processing power to interpret the information and are often more limited in the temporal bandwidth that can be achieved.

The following subsections discuss the types of optical eye trackers most commonly used. These are limbus (reflectivity pattern) trackers, corneal reflection trackers, CR/4PI (dual Purkinjie image) trackers, and CR/pupil trackers. Additional detail can be found in references [2.3-1] and [2.3-25].

2.3.2.3.2. Limbus (Reflectivity pattern) Tracking

Limbus Tracking Technique. The class of eye tracker often referred to as "limbus tracker" uses a small number of solid state photo detectors to measure light reflected from different regions across the front surface of the eye. Typically an IR or near IR LED illuminates the eye, and two or more photo detectors are arranged so that their receptive fields cover the central horizontal axis of the space between the eyelids. Typically, an additional two or more detectors, on the same eye or on the alternate eye, have receptive fields that are arranged vertically. Other systems aim detectors at the horizontal boundary between the eye and the lower eye lid.

Coupling between illuminator and detector is dependent on different reflectivity of different parts of the eye. When making horizontal eye position measurements these devices are influenced primarily by movement of the limbus (boundary between the iris and sclera), which often forms the most dramatic light/dark boundary. Reflectivity also varies between the iris and the pupil, between the eye lids and other parts of the eye, and, to some extent within the iris, sclera, and eye lids. The photo detectors actually respond to the combined effect of this complex pattern. In most cases the number of detectors is too small, and their receptive fields

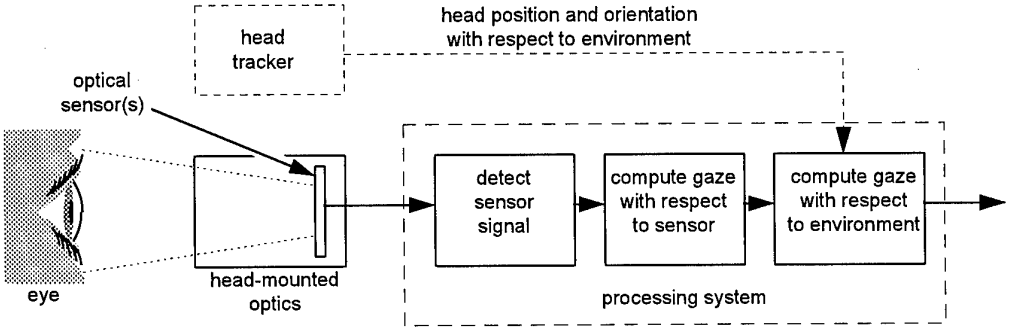


Figure 2.3-4. Schematic showing generic optical gaze tracker

too large to distinguish and track a specific feature.

The top and bottom edges of the limbus are often obscured by the eye lids, and vertical measurements must therefore rely on pupil-iris contrast, and contrast between the eye lids and other parts of the eye, as well as the limbus boundary. The lower eye lid tends to move proportionately to vertical eye pointing direction, and some systems primarily detect the lower eye lid boundary for the vertical measurement.

Most limbus trackers require the sensors to be placed very close to the eye. Often, the light source is modulated for phase sensitive detection in order to enhance the signal to noise ratio. Use of only a small number of analog sensors makes it possible to rely heavily on analog signal processing. A simple schematic for a limbus tracker is shown in Figure 2.3-5.

As with most eye trackers, a calibration scheme must be used to map the relative reflectance signals to a useful line of gaze reference. Movement of the eye in one axis affects the reflectivity pattern in the other (a slightly different

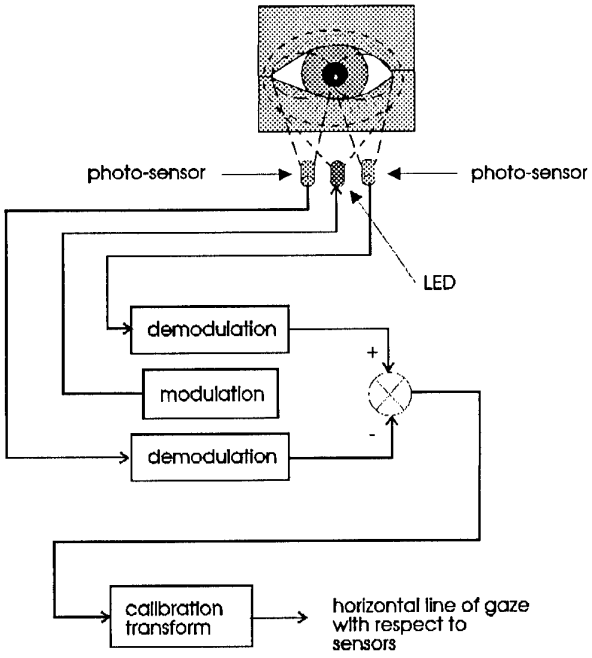


Figure 2.3-5. Schematic showing simple limbus tracker for horizontal measurement

arrangement of features becomes exposed to the sensor receptive fields) often resulting in very significant cross talk effects which must be removed by the calibration mapping algorithm.

There are quite a few developed systems in the limbus tracker category, although they vary substantially in their implementation details.

Limbus Tracking Performance. Limbus trackers usually have very high data update rates and very good resolution; but limited accuracy, precision, and range. Data update rates are often 1000 samples/sec or faster, and eye position resolution may vary from 0.25° to minutes of arc depending on implementation details.

The reflectivity patterns detected by these devices move more or less as would a single landmark on the surface of the eyeball. Shifting of the optics on the head would therefore be expected to cause an error of about 5° for 1 millimeter of movement parallel to plane of the sensor; however, shifting of the optics also perturbs the calibration mapping in complicated and unpredictable ways, making compensation difficult even if the amount of shift can be independently measured.

Accuracy and precision are very hard to quantify because they depend upon the amount of head gear slippage. Errors of less than 1 degree visual angle in the horizontal axis and 2° in the vertical axis are typical over short time periods, but errors of several degrees would not be unusual over longer periods and in the presence of vigorous head motion. Better accuracy is possible when great care is taken to prevent slippage errors.

The response tends to become very non linear and difficult to calibrate over an eye movement range of more than 30° to 40° in either axis. The effective range may be extended somewhat by systems that use more than two sensors for each axis, coupled with digital logic.

Limbus Tracking - Practical problems. Limbus trackers have undependable long term accuracy. The requirements for positioning the sensor array close to the eye tends to produce visual obstructions, creates a potential safety problem, and results in questionable compatibility with other head mounted optics

Limbus tracking - Prognosis. Devices that detect reflectivity patterns, commonly referred to as limbus trackers, are relatively inexpensive tools that have great value in studies of eye movement dynamics in a laboratory setting. Static accuracy limitations and potential mechanical interference problems make them unlikely candidates for use in performing cockpit control tasks.

2.3.2.3.3. Corneal Reflection (CR) Tracking

CR Tracking - Technique. Corneal Reflection (CR) tracking is a single feature technique that follows the mirror image of a light source produced by reflection from the anterior (outer) surface of the cornea. A head mounted source can be used to produce a corneal reflection that appears as small spot that is very bright relative to surrounding features. The first CR trackers worked by using a set of beam splitters and lenses to optically superimpose a CR in the image path of a head mounted movie camera.

More current systems use signal processing to identify the CR on the video image from a head mounted camera, or some other type of solid state sensor. The CR is formed by a head mounted IR light source. A hot mirror (beam splitter that reflects IR and transmits visible wavelengths) is usually used to reflect the eye image to the sensor optics while still allowing unobstructed vision for the wearer.

Because of the relative brightness of the CR it can be discriminated with good reliability using simple intensity threshold techniques. As with other optical techniques, maximum accuracy and linearity require calibration to account for individual eye geometry.

There are few commercially available systems in current use.

Mention must be made of an interesting variation on the basic CR tracker in the form of a "three-dimensional optometer" which has been devised by Takeda *et al* [2.3-26]. A large spherical mirror, a partially reflecting mirror, a relay lens and a pair of servo-controlled gimbaled mirrors have been arranged to provide a view of the eye along its optical axis by dynamically adjusting the gimbaled mirrors so that the CR is centred on a video camera. The angles of the servoed mirrors then relate directly to the eye pointing direction, and the pupil size can also be measured by analysing the video image. The availability of such a frontal view of the eye, over a wide range of eye pointing directions, is exploited by an additional set of infra red optics based on a commercial auto-refractometer to measure the focal state of the eye lens. Although the whole system is large and heavy (~35 kg), it has been mounted on a cantilever with a counterweight so that it can be attached to the subject's head and allow some head movement. Unlike a CR/4PI system, accommodation, pupil size and eye pointing direction are measured simultaneously over a wide range and to good accuracy without needing a mydriatic to dilate the pupil.

CR tracking - Performance. The potential resolution of the measurement is extremely good, but depends upon the specific implementation. Frecher et al [2.3-27] achieved resolutions of under 1 arc minute. Instead of a two dimensional video sensor, they used cylindrical lenses to image a CR onto a pair of orthogonally oriented linear sensor arrays. Since the CR covered more than one pixel with a known intensity distribution, they were able to compute its position to sub pixel resolution. Video camera based systems more commonly achieve resolutions ranging from 0.1° to 0.5°.

As with any single feature technique, accuracy and precision are limited primarily by head-gear slippage. From the central position (eye optical axis pointing directly at illumination source), a 1 mm sensor shift is equivalent to about 10° of eye rotation. Errors of more than 10° would not be unusual in a fighter aircraft, as the combination of rapid voluntary head motion, high-G and vibration cause even a well-fitted helmet to shift transiently by at least 5 mm, and enduring slippage of about this magnitude is also likely [2.3-28] If a bite bar is used to stabilize the head with respect to the sensor optics, precision on the order of arc minutes is probably possible but this is only practical in the laboratory.

Range is limited to approximately ±25° by the CR excursion to the edge of cornea.

Frequency response depends upon the implementation. Frecher *et al* [2.3-27] achieved 1000 sample/sec data update rate with linear sensor arrays. Video based systems are more commonly limited to 50 or 60 samples/sec.

CR tracking - Practical problems. The biggest problem is the high sensitivity to errors induced by sensor shifts. Mechanical challenges created by the need for non-intrusive illuminator and sensor optics are similar to those for the differential CR/Pupil tracking as discussed below.

CR tracking - Prognosis. Large shift-induced errors make the CR technique less attractive than differential pupil-CR tracking

2.3.2.3.4. Differential CR/4PI Tracking

CR/4PI Tracking - Technique. CR/4PI tracking is a dual feature technique that measures the relative positions of the first and fourth Purkinje images. The first Purkinje image (also called corneal reflection or CR) is the mirror reflection from the outer surface of the cornea, and the fourth Purkinje image (4PI) is the reflection from the rear surface of the lens.

The only commercially available CR/4PI system images the two features onto separate quadrant detectors, and uses separated closed-loop servo-controlled pathways to keep the features centered on the detectors. The servo-control signals are a measure of the feature positions, and the horizontal and vertical separation between the features is used to calculate the eye pointing direction.

The outer corneal surface reflects about 4% of the incident illumination and the CR is very bright in comparison with the rest of the eye surface and remains visible over the entire extent of the corneal surface. The rear surface of the lens reflects only about 0.02% so the 4PI is quite dim by comparison and remains visible only within the pupil opening (the area where the lens surfaces are exposed). It should be noted that, as the curvature of the eye lens changes with the accommodation state of the eye, commercial systems simultaneously track the depth of the 4PI image as well as the lateral displacement to measure also the accommodation of the eye.

The lab model of the system uses an IR illumination source and a hot mirror beam splitter to direct IR light to the eye tracker optics while allowing the user an unobstructed view of the forward visual field. AC modulation of the source and synchronous detection are used to enhance the signal to noise ratio.

The optics can be mounted on a servo controlled X-Y-Z platform to allow small head movements, but it has not yet been reduced to a size that would allow head mounting.

CR/4PI Tracking - Performance. Precision and achievable accuracy of the available CR/4PI system are reported to be on the order of 20 arc seconds and 1 arc minute, respectively.

The analog nature of system, permits a very high temporal bandwidth. The available system typically modulates the sensor and light source at 4 KHz, and has an effective temporal bandwidth on the order 500 Hz

The range of CR/4PI tracking is limited to ±10°, extendible to ±15° if the pupil is enlarged by muscle-relaxant drops to reveal lens surfaces at larger angles.

CR/4PI Tracking - Practical problems. The range of measurable eye movement is quite limited, and the only current system is engineered as a large bench-mounted optical assembly which is impractical for airborne application.

CR/4PI Tracking - Prognosis. It would probably be both impractical and prohibitively expensive to develop the CR/4PI device as a head-mountable servo-optical unit. While a valuable laboratory tool because of its exquisite precision and high bandwidth, it is likely to remain unsuitable for airborne application.

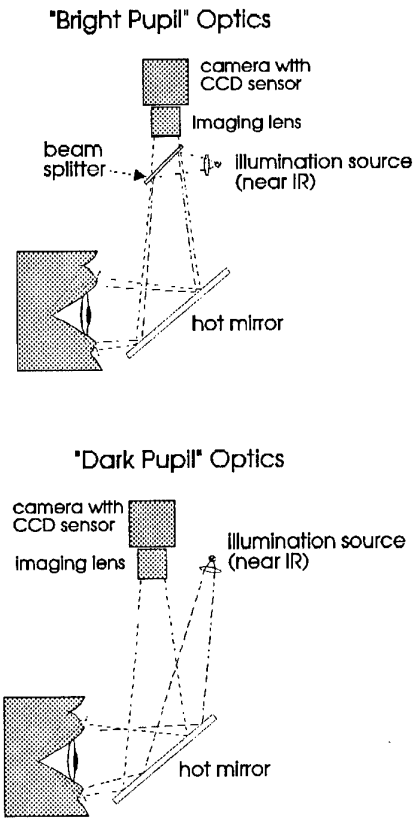


Figure 2.3-6. Typical "bright pupil" and "dark pupil" optics for CR/Pupil tracking.

2.3.2.3.5. Differential CR/Pupil Tracking

CR/Pupil Tracking - Technique. CR/Pupil tracking is dual feature technique that measures the relative positions of the pupil and a corneal reflection. The dual feature principle allows gaze to be measured with respect to either head mounted, or non head mounted optics. In the case of head mounted optics, CR/Pupil systems are reasonably insensitive to headgear slippage.

Generally the eye area is illuminated by a near infra red source (or multiple sources) and a solid state video camera (CCD or CID sensor with related electronics) captures an image of the eye. The camera is typically filtered to receive only light of the wavelength produced by the infra red source (or sources) that are part of the eye tracker.

If the optics (camera, illuminator, and lenses) are mounted to the user's head gear, a hot mirror (beam splitter that reflects IR and transmits visible wavelengths) is usually used to reflect near IR light to the optics while still allowing unobstructed vision for the wearer. This is illustrated is Figure 2.3-6. Alternately, non head mounted optics may use a moving mirror or moving camera platform to follow head motions.

The eye acts as a retro-reflector. If the eye illumination beam is coaxial with the camera, light reflected back from the retina is captured by the camera making the pupil appear to be a bright circle. This accounts for the red eye effect sometimes produced by flash photography. The apparent brightness of a bright pupil image decreases with square of pupil diameter. Off axis illumination produces the more familiar dark (black) pupil image (see Figure 2.3-6).

Dark pupil images provide easier pupil detection in very bright environments (e.g. sunlight), whereas bright pupil images provide easier detection in darker environments.

The corneal reflection is significantly brighter than any other visible feature, including a bright pupil image, and is relatively easy to detect so long as it is not obscured by the eye lids or confused with corneal reflections from some external sources.

Real time image analysis is used to identify the pupil and corneal reflection and find their centers. Relative feature brightness is often a primary discrimination criterion, but more and more sophisticated pattern recognition techniques are being used as the amount of readily available digital processing power increases. This makes it possible to recognize the features of interest in real conditions and cope with extraneous reflections, partial eye lid occlusion and motion-induced blur [2.3-29].

Calibration is required to account for individual eye geometry and optics placement.

There are numerous developed CR/pupil systems, but currently no militarized systems, and no systems robust enough for operational military use

CR/Pupil Tracking - Performance. Range is limited to about ±25° by CR excursion within cornea for systems with a single illumination source, but can be extended considerably for multiple sources, especially on the horizontal axis [2.3-58].

Vertical range is also limited by eye-lid occlusion of the pupil. Eye-lid occlusion is highly variable between subjects, but occlusion by the upper eye-lid is very often a problem when line of gaze is more than 5° below the optical axis of the sensor. The optics are usually set so that the camera axis is depressed to give a view from about 5° below the nominally straight ahead direction.

Range is measured with respect to the sensor. For systems with non head mounted optics this defines a locus of measurable gaze vectors with respect to the cockpit. For systems with head mounted optics it defines only a locus of gaze vectors with respect to the head, and measurable line of gaze with respect to the airframe is limited only by head motion.

Accuracy is usually on the order of 1° visual angle for CR/Pupil systems. It may vary from 0.5° or better during very careful laboratory tests with selected subjects, to > 2.0° under difficult conditions or with difficult subjects. Precision and resolution are usually in the range of 0.1° to 0.5° depending upon the particular system and upon operating conditions.

Frequency response is usually limited by video frame rates to 50 or 60 samples/sec. Higher update rate sensors are available, but require sacrifices in spatial resolution, physical size, and sensitivity.

CR/Pupil Tracking - Practical problems. The need for a non intrusive optics mechanism, and provision for positional

and focus adjustment presents a mechanical challenge for either head mounted or remote optics.

In the area of performance, the greatest difficulties are introduced by bright sunlight, and radical changes in ambient light conditions. Bright pupil systems depend upon the pupil appearing brighter than surrounding features (iris, sclera, eye lids, etc.). Sunlight lights up surrounding features while dimming the bright pupil by causing the pupil to constrict. Because of its intense radiance in all wavelengths, sunlight can also be very difficult for dark pupil systems. Reflections are formed which may obscure the pupil or be confused with the CR, and the dark pupil may actually brighten slightly. Some IR rays that are parallel to a head mounted eye tracker optical path (the section of the path between the hot mirror and eye) do penetrate the hot mirror and reflect back from the retina, producing some bright pupil effect. Additionally, the limited dynamic range of current video cameras is easily exhausted and, despite automatic gain control, sunlight tends to saturate the output and degrade the effective contrast of the CR.

Once features are correctly recognized, the precision of the pupil center computation can be compromised by pupil boundary irregularities and by partial occlusion of the pupil by eyelids, eye lashes, and corneal reflections. The precision of the CR/pupil measurement is also limited by the fact that the pupil center does not remain fixed with respect to the visual axis of the eye as the pupil dilates and constricts, although the amount by which this varies is not well known.

Non head mounted systems face several additional problems. In order to handle large motions of the head with respect to the eye camera, the optics unit must either include a moving mirror, a moving camera platform, use wide angle optics, or use multiple cameras. These requirements severely compromise reliability and are difficult to implement in limited cockpit space. In addition, non head mounted systems are prone to measurement loss when objects (e.g. user's hands and arms) intersect the eye camera optical path, and may have significant problems looking at the eye through other head mounted optics such as head mounted displays, visors, and sunglasses. Specular reflections from these surfaces pose especially difficult problems.

CR/Pupil Tracking - Prognosis. Head mounted optics will probably be required to meet dependability and range requirements, and to allow operation with other head mounted optics.

Multiple illumination sources may be required for adequate range with respect to head gear, and dark pupil or some combination of dark and bright pupil recognition may be required to handle the full range of ambient conditions in the aircraft environment.

Further development is warranted to adequately reject sunlight interference, by suitable filtering, sensor shuttering, illuminator modulation, and/or other optical techniques. Further development is also warranted to increase reliability by using improved feature extraction techniques. Increases in available computation power should allow progressively more sophisticated image recognition and center determination algorithms. Future possibilities range from classical techniques to neural nets.

2.3.2.4. Calibration

All eye tracking methods require a transformation to convert the measured quantity to the desired quantity. For example separation between the pupil and CR must be converted to point of gaze coordinates on a surface, or a line of gaze vector in a particular coordinate frame. For all eye tracking methods discussed, with the possible exception of the scleral coil, the transformation parameters vary between subjects, optics placement, and other conditions.

Calibration refers to a procedure for gathering data, and for using the data to compute transformation parameters. The procedure usually consists of asking the subject to look at number of pre-defined points, while storing samples of the measured quantity (e.g. pupil and CR position).

The transformation can either be a form of interpolation, a set of continuous equations, or some combination of these. The details vary widely among available systems. Theoretically, the transformation can remove any systematic error that is a function of the measured quantities. In practice, there is a limit to the amount of data that it is reasonable to gather with a calibration procedure, and therefore a limit to the complexity of the transformation. Transformation complexity also affects computation time and achievable measurement update rates, but because of the continuing increases in available computation power this is becoming progressively less of a problem.

The accuracy and linearity of eye tracker measurements are limited by the underlying precision (repeatability) of the measured quantity. Up to that limit, accuracy and linearity are determined by the quality of the calibration and transformation scheme. Adding calibration target points, and using the additional data to add interpolation points or to increase the order of a polynomial transform, usually improves accuracy, but with diminishing returns. Typical calibration schemes require 5 or 9 pre-defined points, and rarely use more than 20 points.

An example of a 2 dimensional interpolation scheme can be found in references [2.3-30] and [2.3-31]. A cascaded polynomial curve fit method is described in reference [2.3-32]. Many other variations of these schemes are possible.

2.3.2.5. Fixation Filtering

Scanning behavior is described by a series of fixations, saccades, and smooth pursuits. When using gaze measurement to determine point of regard, as often required by gaze based control techniques, it may be desirable to recognize fixations while filtering out saccades, pursuits, and measurement noise. This is usually done by looking for periods of longer than some threshold time during which gaze remains within a threshold area or during which eye movement velocity remains below a threshold. Typically, the threshold time period is on the order of 100 msec, and the threshold gaze change is on the order of 1deg visual angle. If velocity thresholds are used, the value is typically about 10 deg/sec. There is no precise physiological definition of a fixation, and fixation algorithm parameters can vary substantially since they must be chosen to best accommodate the noisiness and other performance characteristics of the gaze measurement device as well as the requirements of the particular task at hand.

Off line fixation detection, with the luxury of looking backward in time, can be more precise than on line filtering.

For example, once it is known by an on line algorithm that gaze has been fixed for the requisite time and therefore a fixation is in progress, the fixation start time has already passed. Off-line analysis, on the other hand, can make use of the more precisely determined start point.

Examples of specific algorithms, which are generally far more involved in their details than the general description given above, can be found in references [2.3-33], [2.3-34], and [2.3-35].

2.3.2.6. Safety

General safety issues associated with helmet mounted equipment are discussed under *Some Proposed Applications of Alternative Controls* (section 4.1). These requirements involve quick disconnect features, analysis of inertial force effects under applicable G loading, and, for certain types of aircraft, behavior during emergency ejection. Placing any object, particularly a set of frangible optics, close to the eye would not be acceptable, since the eyes could easily be damaged in a minor frontal impact.

Most optical eye trackers illuminate the eye with a near infra red source, and care must be taken that such illumination is not harmful to the eye. The primary considerations for infra red (including near infra red) are potential thermal injury to the retina, lens, and cornea. Photochemical injury to these structures is also a consideration for ultraviolet and, in the case of the retina, blue light; but eye trackers do not typically use these wavelengths.

The most applicable recommendations for human exposure to non laser optical sources are "Threshold Limit Values (TLVS) published *by The American Conference of Governmental Industrial Hygienist (ACGIH)*. Standards for laser sources, which are often used to evaluate non laser sources as well, are published by several organizations including the American *National Standards Institute (ANSI)*, The Federal Food and Drug Administration (FDA), and *The International Electrotechnical Commission (IEC)*. The most recent ANSI standard *is Standard Z136.1 (1993), Safe Use of Lasers*. The applicable FDA standard is Title 21 Code of Federal Regulations Part 1040 (21CFR1040). The current IEC standard is Publication 825-1 (1993), which has been expanded to include LEDs as well as laser sources. The standards vary, with the IEC standard currently being the most restrictive. Under IEC standards, for example, the source must be safe even if viewed from the closest mechanically possible distance through a magnifying glass. The standards require evaluations based on power, wavelength spectrum, divergence of the beam, apparent angular subtense of the source as viewed by the subject, expected exposure duration, angle of exposure, etc. An excellent review of light source safety can be found in Sliney and Wolbarsht [2.3-36].

Eye tracker illumination sources should be carefully evaluated according to the most applicable standard by suitably trained professionals. The following guidelines for a particular restrictive case are from ACGIH recommendations, and are presented only as an example.

To protect the cornea and lens, long (> 1,000 second) exposure to non laser infrared radiation ($770\,nm < \lambda < 3\mu m$) should be limited to irradiance at the cornea of no more than 10 mW/cm^2. To protect the retina, total radiance (W/cm^{2}sr), of a non laser near infra red source ($770\,nm < \lambda < 1440\,nm$), where a strong visual stimulus is absent, should be limited to $\frac{0.6}{\alpha}$, where α is the angular subtense, in radians, of the source as viewed by the observer.

2.3.3. APPLICATIONS TO DATE

2.3.3.1. Eye Designation

Eye designation has been investigated in the lab, and generally has been found to be as fast or faster than manual designation so long as the eye tracker is working dependably and so long as the task employs large enough targets to be well within the accuracy capability of the eye tracker. Applied use of eye designation has been primarily restricted to systems that facilitate communication for people with motor control disabilities, and there are several commercially available systems that specifically support this application.

Performance of eye designation tasks may be sometimes be significantly enhanced by use of fixation filtering algorithms (see section 2.2.2.4), but in general, accuracy of unobtrusive eye tracking systems does not permit as fine a control capability as mouse, trackball, or other manual techniques.

Calhoun and Janson [2.3-37], and Calhoun et.al. [2.3-38] compared time required to select switches manually with selection times for eye position control followed by button press confirmation. They found that selection times were very similar. Calhoun and Janson [2.3-39] found eye control to be significantly faster than head control for target selection and weapon selection tasks.

Ware and Mikaelian [2.3-40] used an eye tracker with non head mounted optics to control a display cursor. Subjects were required to move a cursor to a highlight a rectangle, and confirm the selection with one of several methods. Confirmation methods included a button press, and a 400 msec on-target dwell time. Subject performance was plotted in terms of the Welford formulation of Fitts law. The requirement to keep targets large compared to eye tracker accuracy resulted in a relatively small range of index of difficulty values towards the low end of the difficulty scale, but within this range performance was faster than selection with a mouse.

Borah [2.3-41] tested a task similar to the Ware and Mikailian task using a head mounted eye tracker coupled with a magnetic head tracker. Subjects were required to designate circular targets of different sizes appearing at random positions on a video screen, and confirm the selection with a button press. No point of gaze cursor was displayed to the subject, but the target changed color when the measured point of gaze was within the target boundary. Regression results were very similar to Ware and Mikaelian.

Jacob [2.3-35] used an eye tracker with non head mounted optics as part of a system that allowed users to position objects on a display, operate pull down menus, and display expanded information windows relating to specific objects. He used a fixation filter to help stabilize the measurement, and used varying techniques to confirm actions depending on the consequences of inadvertent action. Actions that were benign and easily reversible required only short fixations for activation. Actions that were not as easily reversible required longer fixations or manual confirmations. He found that when the eye tracker was working accurately and dependably

it felt as though the system were reading the user's mind, but when eye tracker performance was not stable enough or not accurate enough it was extremely frustrating to the user.

Communication applications for the handicapped have generally taken the form of eye controlled computer input devices, for people without significant motor function (other than eye movement) or speech capability. For this application cost must remain relatively low, and performance requirements are relative to the very limited set of alternate techniques available. A small number of companies make devices targeted especially for this application, and several experimental systems have been built that have not yet led to commercial devices.

2.3.3.2. Eye tracking under High G and Vibration

It seems reasonable that high G levels should have far less affect on a person's capability to make eye movements than on other types of body movement. The eye is well supported, and has a relatively small moment of inertia. Because of its roughly spherical, homogeneous structure inertial forces would not be expected to produce large rotational moments. Furthermore, eye movements do not cause the disorienting motion sensations that can be produced by moving the head, and hence the vestibular sensors, in the presence of high inertial forces. The limbs and head begin to feel extremely heavy at levels above 3 Gz. No visual scanning difficulties are reported in the literature for centrifuge studies even at Gz levels that discourage head movement, although, with the exception of Sandor (2.3-40), eye movements were generally not explicitly studied.

The affect of high Gz on vision itself (as opposed to movement of the eyes) is well documented. Peripheral vision begins to fade ("vision tunneling") at levels of 3-4.5 Gz, followed by complete vision loss ("black out") at levels of 4-5.5 Gz, and unconciousness at levels between 4.5 and 6 Gz [2.3-42]. These levels vary between individuals and can be extended by mechanical "G suit" devices, physical fitness, and training [2.3-42].

Sandor, et al [2.3-43] successfully measured eye and head movements to determine point of gaze during a visual tracking task under loads of up to 5 Gz . Eye movements were measured using the pupil-to-corneal reflection method, and head position was measured with an electro-optical system. Subjects tracked targets which moved ±20 degrees in azimuth and ±10 degrees in elevation, using both eye and head movement, without reported difficulty.

Eye trackers have been used successfully, for research purposes, in a variety of ground vehicles, but systematic studies of eye movement measurement under varying levels of vibration do not seem to be available. The transmission of vibration from the seat to the head of an operator has however been studied extensively [2.3-44, 2.3-45, 2.3-46, 2.3-47], and is discussed in more detail in section 2.2.1.2 (under *Head-Based Control*). Head pointing is disturbed significantly by whole-body vibration, especially in the 3 to 6 Hz region. Fortunately the balance organs in the head, the semi-circular canals and the otoliths of the vestibular system, are linked closely with the eye muscles to stabilise the eyes against such motions. As this vestibulo-ocular reflex (VOR) is effective up to about 10 Hz [2.3-48] the human ability to point with the eye should be relatively immune to disturbance by whole body vibration, although the eye direction measurement system may be adversely affected. Allison, Eizenman and Cheung [2.3-58] have however been able to use a fast, accurate head-mounted CR/pupil eye tracking system, in conjunction with a magnetic head tracker, to study abnormality of the VOR by assessing the velocity of compensatory eye movements as the subject attempts to maintain fixation on a nearby target while his head is turned rapidly through a small angle by the experimenter. Normally the eye angular velocity is equal and opposite to that of the head, a VOR gain of unity, but for subjects with damaged vestibular nerves the initial compensatory eye movements were very small, corresponding to a VOR gain of about 0.35.

2.3.3.3. Eye tracking in Conjunction with other Control Techniques

Eye tracking has been shown to be useful as a provider of contextual as well as positioning information in conjunction with other control techniques.

Starker and Bolt [2.3-49] tested a method for intelligent interface with a display based on eye movement pattern analysis.

Glenn et al [2.3-50] implemented an integrated eye tracking and voice recognition system for user interaction with a graphical display. Gaze information is used essentially as a cursor control with specific time tagged words (e.g. "NOW") acting as confirmation switches to indicate when the gaze information is to be acted upon.

Using an extension of the GOMS task analysis model [2.3-51, 2.3-52] as a design tool, Hatfield [2.3-53] has demonstrated an eye/voice user interface technique for a military aviation mission planning task. Point of gaze information is used primary to provide context and position information for verbal commands. Gaze tracker data is fixation filtered to form time stamped fixation events. Speech is also detected, parsed and time stamped. When verbal "referents" are detected they are matched with corresponding fixation positions. For example the command "nav designate steer point" sets the steer point to the radar display position being fixated at the time the verbal referent "designate" was detected. In this way some operations that would be sequential with a single control modality can be made concurrent when the modalities are combined.

Since the early 1980's the MIT media lab has pursued the combined use of voice recognition, eye tracking, and gesture (hand position) tracking for user display interaction [2.3-54, 2.3-55]. Information from all three modalities must be individually processed and time tagged, and related to discrete interaction events. The prototype system described by Koons [2.3-55] uses both gesture and gaze data primarily as context information to help disambiguate verbal instructions

Jacob [2.3-35] used eye point of gaze measurement for contextual and positioning information in combination with mouse and voice control. Jacob emphasizes use of the modalities in a natural way that does not force users to learn new behaviors.

Although eye tracking is not significantly addressed, Kocian and Task [2.3-16] present a thorough review of visually coupled systems in military aircraft with emphasis on display and head tracking components. Such visually coupled

Table 2.3- I Summary of Major Eye Tracking Techniques

Method	Typical Applications	Typical Attributes	Typical Reference Frame(s)	Typical Performance
• EOG	• Dynamics of saccades • smooth pursuit • nystagmus	• High bandwidth • Eyes can be closed • In expensive • Drift problem (poor position accuracy) • Requires skin electrodes - otherwise unobtrusive	• Head	• static accuracy: ~3°-7° • resolution: with low pass filtering & periodic re-zero, virtually infinite • bandwidth: ~100 Hz
• Limbus (reflectivity pattern)	• Dynamics of saccades, smooth pursuit, nystagmus • Point of gaze • Scan path	• High bandwidth • Inexpensive • Poor vertical accuracy • Obtrusive (sensors close to eye) • Head gear slip errors	• Head gear	• accuracy: varies • resolution: 0.1° (much better res. possible) • range: ~30° • update rate: 1000 samples/sec
• CR	• Point of gaze • Scan path	• Large head gear slip errors • Unobtrusive • Low bandwidth	• Head gear	• accuracy: 1°-10° • resolution: 0.25° (much better res. possible) • hor. range: ~50° • vert. range: ~40° • update rate: variable
• CR/Pupil	• Point of gaze • Scan path	• Minimal head gear slip error • Unobtrusive • Low bandwidth • Problems with sunlight	• Head gear • Room (airframe)	• accuracy: ~1° • resolution: ~ 0.2° • hor. range:~50° • vert. range: ~40° • update rate: 50 or 60 samples/sec
• CR/4PI	• Dynamics of saccades, smooth pursuit, nystagmus • Miniature eye movements • Point of gaze • Scan path • Image stabilization on retina • Accommodation	• Very high accuracy and precision • High bandwidth • Obtrusive (large optics package, restricted head motion) • Limited range	• Room	• accuracy: min. of arc • range: ~20° • update rate: 500 samples/sec
• Scleral Coil	• Dynamics of saccades, smooth pursuit, nystagmus • Miniature eye movements • Point of gaze • Scan path	• Very high accuracy and precision • Invasive • Very obtrusive	• Room	• accuracy:~15 sec arc • resolution: ~1arc min. • range: ~ 30° • bandwidth: ~200 Hz

systems are likely to be major components of any aircraft control or interaction system involving eye trackers.

Eye tracking has also been used in the laboratory as one component of a multimode positioning control. Borah [2.3-41] tested a modality switch between eye control for gross positioning and head control for fine adjustments. The technique enabled selection of targets that were smaller than eye tracker accuracy, but the task required some concentration and needs refinement to be practical. Manual, rather then head position, control may be preferable as the fine control mode.

2.3.3.4. Use of Eye Position Feedback

Eye position feedback allows subjects to view their own measured point of gaze on the scene surface. It might take the form of a cursor, superimposed on a terminal screen being viewed by the subject, and always positioned at the point of gaze measured by some gaze tracking device. If the gaze tracker were perfect such a cursor would always be imaged at the same spot in the center of the fovea, and in fact would seem to the subject to disappear due to retinal accommodation. In actual practice, there is usually some varying offset from the foveal center due to gaze tracking inaccuracy, measurement noise, and latency.

Peli and Zeevi [2.3-56] determined that point of gaze feedback improves smooth pursuit performance. Kenyon, et. al [2.3-57] found that smooth pursuit in the absence of a target is more easily learned and better performed when point of gaze feed back is provided. The benefit degenerates when delay in the feedback is more than 50 msec and when errors become larger than the fovea.

Calhoun and Janson [2.3-39], Jacob [2.3-35], and Borah [2.3-41] all found designation tasks easier when point of gaze feedback is not provided. Note that these studies used video based instruments with less precision and greater latency than the Peli and Zeevi [2.3-56] and Kenyon, et al [2.3-57] research. Also contradicting the findings that point of gaze feedback is disadvantageous, are some unpublished reports that subjects, using a video based gaze tracking device with

similar performance characteristics, were easily able to correct small gaze tracking errors by adjusting their gaze to place the feedback cursor on target. It is likely that tolerance of errors varies between individuals, tasks and conditions. For instance, it would be reasonable to expect a laboratory game-player to accommodate shortcomings which would be deemed totally unsatisfactory by a combat pilot.

2.3.4. REQUIRED ENHANCEMENTS AND PROGNOSIS

Eye tracking is a relatively mature technology only in the R&D domain. No currently available eye tracking systems are dependable enough or automatic enough for operational flight applications, nor are there any current systems available in a militarized configuration. All current devices require a skilled equipment operator (other than the person being measured) for optimal use.

Scleral coil and differential PI tracking seem likely to remain laboratory techniques unless some major leaps in the technology occur. These techniques are by far the most accurate of the major techniques in use, but they both present major practical problems. Scleral coil tracking is invasive and requires a Helmholz coil that will probably be difficult to integrate on aircrew helmets. Differential PI tracking has too restrictive a range limitation and too complex an optics package to be easily helmet mounted and ruggedized.

EOG may very well have a place in aircraft as a back up measurement system, an enhancement to add temporal bandwidth to another type of eye tracker, or for use in some control function that requires only detection of eye movement, rather than absolute line of gaze. The accuracy of EOG alone is never likely to be adequate for line of gaze determination.

Differential CR/pupil tracking systems are generally the most unobtrusive eye tracking systems available, and, with head mounted optics, currently come closest to being appropriate for operational use in flight. Those systems, using dark pupil optics along with some form of illuminator strobing and sensor shuttering, are currently best able to operate in daylight and under vibration. The static accuracy (about 1° visual angle) and range (50° horizontal and 40° vertical field with single illumination source) of current CR/pupil tracking devices is adequate for implementing or assisting a variety of tasks in the aerospace environment, although increased accuracy would certainly expand the potential use for eye tracking.

No currently available CR/pupil systems are yet nearly robust enough, automatic enough, or properly integrated with aircrew head gear and military electronics. Robustness must be significantly improved to ensure dependable operation for different users under varying light conditions in operational environments. Automatic operation must be significantly enhanced so that the user can don the equipment and calibrate the system without help, and there-after depend upon proper operation with no intervention by a second person. Optics must be integrated with the appropriate head gear and head mounted display systems, and both optics and electronics must be hardened and militarized. Work is underway in all of these areas, and there is no reason to think that such enhancements cannot be realized with currently available optics, sensor, and processing technology.

Significant improvement of CR/pupil system accuracy is clearly possible, but far less certain, especially in operational environments. Improvements can theoretically be made by using increased processing power to more accurately find the center of the oval pupil in the presence of partial image occlusions, and ragged edges; use of higher order calibration schemes to remove more of the systematic error; use of additional variables in calibration, such as pupil diameter, to further account for systematic effects; use of precision sensor arrays, or sensor arrays that are mapped and compensated for spatial non-linearities; etc. Such gains may be more than counter-balanced, however, by additional error introduced under the rigors of operational environments. Furthermore, the lengthy, careful calibration procedures probably required for exquisite accuracy may be contrary to operational imperatives for quick and easy set-up. Designers may want to consider eye tracking tasks tailored to require less rather than more accuracy and precision in order to improve the chances of meeting robustness and ease of use requirements.

Major eye tracking techniques are summarized in Table 2.3-I.

2.3.5. REFERENCES

2.3-1 Young L. R. and Sheena D., "Eye Movement Measurement Techniques", in Webster (Ed) "Encyclopedia of Medical Devices and Instrumentation", New York, John Wiley & Sons, 1988

2.3-2 Borah, J., "Helmet Mounted Eye Tracking for Virtual Panoramic Display Systems - Volume II: Eye Tracker Specification and Design Approach", US Air Force report AAMRL-TR-89019, August 1989.

2.3-3 Hallet,P.E., "Eye Movements", in Boff, K. R., Kaufman, L., and Thomas, M. P. (Eds) "Handbook of Perception and Human Performance - Vol. 1", New York, John Wiley and Sons. 1986.

2.3-4 Young, L.R., "The sampled data model and foveal dead zone for saccades", in Zuber B. L. (Ed) "Models of Oculomotor Behavior and Control", Boca Raton, CRC Press, 1981.

2.3-5 Julez, B., Gilbert, E. N., Shepp, L. A., and Risch, H. L., "Inability of humans to discriminate between visual textures that agree in second-order statistics - revisited", Perception, 2, 1973, pp 391-404.

2.3-6 Bergin, J. R. and Julez, B., "Parallel versus serial processing in rapid pattern discrimination" Nature, 303, pp 696-698.

2.3-7 Scinto, L. F. M., "Retinal inhomogeneity and the allocation of focal attention during fixation", in "The Annual Meeting of the Applied Vision Association" St. John's College, Oxford. 1988.

2.3-8 Yarbus A. L., "Eye Movements and Vision", New York, Plenum Press, 1967.

2.3-9 St.Cyr G. L. and Fender, D. H., "Non-linearities of the human oculomotor system: Gain", Vision Research, 9, 1969, pp 1235-1246.

2.3-10 Michael, J. A., Melvill Jones, G., "Dependence of visual tracking capability upon stimulus

2.3-12 Steinman, R. M. and Collewijn, H., "Binocular Retinal Image Motion During Active Head Rotation", Vision Research, 20, 1980, pp 415-429.

2.3-13 Yamada. M., "Head and eye coordination analysis and a new gaze analyzer developed for this purpose", in d'Ydewalle G. and Van Rensbergen, J. (Eds) "Visual and oculomotor functions", Elsevier, 1994, pp 423-434.

2.3-14 Bahill, A. T., Adler, D. and Stark, L. "Most naturally occurring human saccades have magnitudes of 15 degrees or less", Investigative Ophthalmology, 14, 1975, pp 468-469.

2.3-15 Co-ordinate systems for describing eye movements. Section 1.903 in Boff, K. R. and Lincoln J. E., (Eds) "Engineering Data Compendium, Human Perception and Performance", US Air Force A.A.M.R.L., Ohio 1988.

2.3-16 Kocian, D. F. and Task, H. L., "Visually Coupled Systems Hardware and the Human Interface", in Barfield, W., and Furness, T. A., (Eds) "Virtual Environments and Advanced Interface Design", New York, Oxford University Press, 1995.

2.3-17 Shackel B., "Eye movement recording by electro-oculography", in "A manual of Psychophysiological methods", North-Holland Publ. Co., 1967.

2.3-18 Viveash J. P., Belyavin A. J., "Eye movements under operational conditions", in Waters M. and Stott J. R. R. (Eds) "Journal of Defence Science", 1, 2, 1996.

2.3-19 Orschansky J., "Eine methode die augenbewegungen direkt zuuntersuchen (opthalmographie)", Zbl. Physiol., 12, 785, 1898.

2.3-20 Delabarre E. B., "A method for recording eye-movements", Amer. J. Psychol., 9, 572, 1898.

2.3-21 Marx E. and Trendelenburg W., "Uber die genauigkeit der einstellung des auge biem fixieren", Z. Sinnesphysiol. 45, 1911, pp 87-102.

2.3-22 Robinson D. A., "A method for measuring eye movement using a scleral search coil in a magnetic field", IEEE Transactions on Biomedical Electronics, BME-10, 1963, pp 137-145.

2.3-23 Ferman L., Collewijn H., Jansen T. C. and Van den Berg A. V., "Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral coil technique", Vision research, 27, 1987. pp 818-828

2.3-24 Brennan, D. H., "Vision and visual protection in fast jet aircraft", in "Visual effects in the high performance aircraft cockpit", AGARD LS-156, 1988

2.3-25 Borah, J., "Helmet Mounted Eye Tracking for Virtual Panoramic Display Systems - Volume I: Review of Current Eye Movement Measurement Technology", US Air Force report AAMRL-TR-89019, 1989

2.3-26 Takeda, T., Fukui, Y., Ikeda, K. and Iide, T., "Three-dimensional optometer III", Applied Optics, 32, 22, 1993, pp 4155-68.

2.3-27 Frecher, R. C., Eizenman, M., and Hallet, P. E., "High precision real-time measurement of eye position using the first Purkinje image" in Gale, A.G. and Johmson, F., (Eds) "Theoretical and applied Aspects of Eye Movement Research", North-Holland, Elsevier Science Publishers B. V., 1984.

2.3-28 Jarrett, D. N., "Helmet-mounted devices in low flying high speed aircraft", AGARD CPP 267, 1979.

2.3-29 Brinicombe, A. M., Boyce, J. F. and Durnell, L., "Direction of regard determination", in Delogne, P., (Ed) "Proc. Intl. Conf. on Image Processing", Lausanne, Switzerland. IEEE Signal Processing Society, 1996.

2.3-30 McConkie, G. W., "Evaluating and reporting data quality in eye movement research", Behavior Research Methods & Instrumentation, 13, 2, 1981, pp 97-106.

2.3-31 Kliegle, R. and Olson, R. K., "Reduction and calibration of eye monitor data", Behavior Research Methods & Instrumentation, 13, 2, 1981, pp 107-111.

2.3-32 Sheena, D. and Borah, J., "Compensation for some second order effects to improve eye position measurements", in Fisher, D. F., Monty, R. A., and Senders J. W., (Eds) "Eye Movements: Cognition and Visual Perception", Hillsdale, Lawrence Erlbaum Associates, 1981.

2.3-33 Lambert, R. H., Monty, R. A. and Hall, R. J., "High-speed processing and unobtrusive monitoring of eye movements", Behavior Research Methods and Instrumentation, 6, 1974, pp 525-530.

2.3-34 Nodine, C. F., Kundel, H. L., Toto, L. C., Krupinski, E. A., "Recording and analysing eye-position data using a microcomputer workstation", Behavior Research Methods, Instruments & Computers, 24, 1992, pp 475-485.

2.3-35 Jacob, R. K., "Eye Tracking in Advanced Interface Design", in Barfield, W. and Furness, T. A. (Eds.) "Virtual Environments and Advanced Interface Design", New York, Oxford University Press, 1995.

2.3-36 Sliney, D. H. and Wolbarsht, M., "Safety with Lasers and Other Optical Sources: A Comprehensive Handbook", Hew York, Plenum Pres, 1980.

2.3-37 Calhoun, G. L. and Janson, W. P., "Eye Line-of-Sight Control Compared to Manual Selection of Discrete Switches", US Air Force report AL-TR-1991-0015, NTIS: AD-A273 019, 1991.

2.3-38 Calhoun, G. L., Janson, W. P. and Arbak, C. J., "Use of eye control to select switches", in "Proceedings of the Human Factors Society - 30th Annual Meeting", 1986, pp 154-158.

2.3-37 Calhoun, G. L. and Janson, W. P., "Eye Line-of-Sight Control Compared to Manual Selection of Discrete Switches", US Air Force report AL-TR-1991-0015, NTIS: AD-A273 019, 1991.

2.3-38 Calhoun, G. L., Janson, W. P. and Arbak, C. J., "Use of eye control to select switches", in "Proceedings of the Human Factors Society - 30th Annual Meeting", 1986, pp 154-158.

2.3-39 Calhoun, G. L. and Janson, W. P., "Eye control interface considerations for aircrew station design", in "Sixth European Conference on Eye Movements", Leuven, Belgium, 1991.

2.3-40 Ware, C. and Mikaelian, H. T., "An evaluation of an eye tracker as a device for computer input", in by Carroll, J.M. and Tanner, P. P., (Eds) "Proceedings of Human Factors in Computing Systems and Graphics Interface Conference", Toronto, Canada. 1987, pp. 183-188.

2.3-41 Borah, J., "Investigation of Eye and Head Controlled Cursor Positioning Techniques", US Air Force report AL/CF-SR-1995-0018, September 1995.

2.3-42 Fong, K. L., "Maximizing +Gz Tolerance in Pilots of High Performance Combat Aircraft, Interim Report", US Air Force report AL-SR-1993-0001, December 1992.

2.3-43 Sandor, P. B., Hortolland, I., Poux, F., and Leger, A., "Orientation du regard sous facteur de Charge Aspects methodologiques Resultats preliminaires", in "AGARD Meeting on Virtual Interfaces: Research and Applications", October, 1993.

2.3-44 Rowlands, G. F., "The transmission of vertical vibration to the head and shoulders of seated men", Royal Aircraft Establishment Technical Report TR-77068, 1977.

2.3-45 Griffin, M. J., "Vertical vibration of seated subjects: Effects of posture, vibration level and frequency", Aviation, Space and Environmental medicine, 46, 1975, pp269-276.

2.3-46 Wells, M. J. and Griffin, M. J., "A review and investigation of aiming and tracking performance with head-mounted sights", IEEE Trans on Systems, Man and Cybernetics, T-SMC/17, 2, 1987, p12094.

2.3-47 Tatham, N. O., "The effects of turbulence on helmet-mounted sight accuracies", AGARD CPP 267, 1979.

2.3-48 Barnes, G. R., Benson, A. J. and Prior, A. R. J., "Visual-vestibular interaction in the control of eye movement", Aviation, Space and Environmental Medicine, 49, 1978, pp557-564.

2.3-49 Starker, I., and Bolt, R. A., "A gaze-response self-disclosing display", in "Proceeding of the ACM CHI '90 Human Factors in Computing Systems Conference", New York, Addison Wesley/ACM Press, 1990, pp 3-9.

2.3-50 Glenn, F. A. Harrington, N., Iavecchia, H. P., and Stokes, J., "An Oculometer and Automated Speech Interface System", in "Analytics, Technical Report 1920", Analytics, Willow Grove, PA, May 1984.

2.3-51 Card, S., Moran, T. and Newell, A., "The psychology of Human Computer Interaction", Hillsdale, Lawrence Erlbaum Associates, 1983.

2.3-52 John, B. E. and Kieras, D. E., "The GOMS Family of Analysis Techniques: Tools for Design and Evaluation", CMU Technical Report CMU-CS-94-181, Carnegie-Mellon University, August 1994.

2.3-53 Hatfield, F., Jenkins, E. and Jennings, M. W., "Eye/Voice Mission Planning Interface (EVMPI)", US Air Force report TR-J103-1, 1995.

2.3-54 Bolt, R. A., "The Human Interface: Where People and Computers Meet", Lifetime Learning Publications, London, UK. 1984.

2.3-55 Koons, D. B., Sparrell, C. J. and Thorisson, K. R., "Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures," in Maybury, M. T. (Ed.) "Intelligent Multimedia Interfaces", Menlo Park, AAI Press/The MIT Press, 1993.

2.3-56 Peli, E. and Zeevi, Y. Y., "Multiple visual feedback loops in eye movement control". in "XII International Conference on Medical and Biological Engineering", Jerusalem, 1979.

2.3-57 Kenyon, R. V., Zeevi, Y.Y., Wetzel, P. A., and Young, L. R., "Eye movement in response to single and multiple targets", US Air Force report AFHRL-TR-84-29, 1985.

2.3-58 Allison, R. S., Eizenman, M. and Cheung, B. S. K., "Combined head and eye tracking system for dynamic testing of the vestibular system", IEEE Trans on Biomedical Engineering, Vol. 43, No 11, November 1996

2.4 GESTURE-BASED CONTROL

2.4.1 INTENTION OF THE TECHNOLOGY

2.4.1.1 Relevance

Most currently available interfaces only make use of discrete pieces of data produced by the user's gestures. This sometimes stems from the very input devices used, such as keyboards, which are intrinsically discrete devices; but even with continuous input devices, such as mice, most of the time only specific events and data points (such as the co-ordinates of the pointer when the user clicks) are taken into account. The trajectory of the mouse during the interaction has little, if any, effect on the result.

Gesture-based interaction aims at taking advantage of the continuity and dynamics of the user's movements, instead of only drawing discrete information from these movements.

This interaction modality has recently begun to appear in devices available to the general public, such as personal digital assistants, but it is still exploited in a very primitive way.

Related issues include person location, face recognition and capture of unintended movement, to the aim of e.g. situation assessment; we will deal briefly with these issues and focus on intentional gesture as a control modality. We will therefore mostly consider hand and arm gestures.

The static control mechanisms provided by finger/hand position or finger direction can be considered as a subset of gesture-based control. In these cases the dynamic portion of the gesture is not needed; pointing to an object requires finger position and direction, but does not need the history of the movement. Similarly, the dynamic gesture of tracing a route on a map can be broken down into a series of static position measurements of the finger-tip. These static methods will also be discussed where relevant.

2.4.1.2 Human Gesture Capabilities and Limitations

2.4.1.2.1 Capabilities

Gesture is a very natural human communication capability; it should thus lend itself to easily learnt (for instance by example) interaction techniques. A distinguishing feature of the gestural communication channel is that it allows one to act on one's environment as well as to retrieve information from it. Three complementary and inter-dependent functions of gesture are pointed out in [2.4-1]:

- the epistemic function, which corresponds to perception. This includes:
  - haptic sense, which combines tactile sense (touch) and kinesthetic sense (awareness of the position of the body and limbs), and gives information about size, shape and orientation,
  - and proprioceptive sense, which informs on weight and movement through joints sensors;
- the ergotic function, which corresponds to actions applied to objects;
- the semiotic function, which is about communication. Examples include sign language and gesture accompanying speech.

We will be concerned with the means of action and expression, thus ergotic and semiotic functions, but also with feedback through the epistemic function.

Typical gesture commands are terse and powerful: a single gesture can encompass a command as well as its arguments. Taking into account the user's movements in all their continuity and dynamics can thus provide more information than current interfaces do and enrich the interaction. For instance, in a drawing program, a linear trajectory can be interpreted as a line-drawing command, while a curved trajectory would start the drawing of a circle. More abstractly, a cross drawn on an object can be a request for deletion; this would be an iconic use of gesture. Even further, provided adequate devices are used, three-dimensional trajectories and the postures of the limbs can be considered, allowing gestures to be recognised more precisely and making direct gestural interaction possible. This can provide for a more natural use of the controlled system and a lower cognitive cost. As a matter of fact, the hand can, to the user, become the very input device used.

The preceding considerations only apply for intentional gestures. Some gestures, such as the movements of the lips that are part of the act of speech, or facial expressions, are generally not deliberate and will most of the time either provide contextual information or be interpreted jointly with another communication means. Spontaneous gestures accompanying speech do not per se constitute a language, but work has been done on typologies; for instance, gestures can stress specific words or sentences, indicate an object or place (deictic gestures), sketch a shape or picture, and so on.

2.4.1.2.2 Limitations

The body movements involved in gestural communication are a source of fatigue; it is thus of prime importance to design the system for the use of concise and not too contorted gestures. Good precision cannot be relied on over time; and as is the case with gaze, it is very difficult for a human to keep a steady position for a long time.

Gesture is made more difficult in a non-benign environment; as with head-based control, hand movements are impaired by G forces and by vibration. So [2.4-2] investigated the transmission of vertical seat vibration to the outstretched hand at frequencies up to 10 Hz, and found involuntary hand motion in both the vertical (pitch) and lateral (yaw) directions. The vertical disturbance showed a resonance peak for the hand at about 2 Hz, whilst hand motion in the lateral axis rose gradually to about 5 Hz, beyond which it had a fairly flat response.

While the kinesthetic sense gives one an indication of the positions of the body and limbs, it is not sufficient to ensure that commands are adequately taken into account; hence the need for adequate feedback.

Gesture presents large intra- as well as inter-subject variability. The impossibility of reproducing exactly the same gesture twice is a source of potential precision and recognition problems. Differences between individuals also means some training of the recognition systems is generally needed.

Another problem in free gesture recognition is similar to one encountered in free speech understanding: a continuous stream of position data is received and has to be converted into a series of gestures considered as lexical entities; a further complication is the fact that co-articulation of gestures modifies the individual gestures, as is the case with phonemes. This leads to the problems of defining and recognising the beginning and ending points of a gesture. A number of systems avoid these difficulties entirely by limiting recognition to static postures.

Still another case for concern is the immersion problem, especially in the case of unobtrusive methods of gesture capture: if every movement can be subject to interpretation, the user will be deprived of external communication when using the system for fear that a movement could be mistakenly interpreted. The only solution is to provide the system with an effective and unobtrusive way of detecting whether a gesture is indeed addressed to it.

2.4.2 OVERVIEW OF APPROACHES

2.4.2.1 Intentional Gesture

2.4.2.1.1 Devices Overview

Human gesture can be captured using a variety of hardware devices. Contact devices, besides classical ones such as mice, trackballs, trackpads and touch screens, include a variety of more exotic ones such as spaceballs, 3-D mice, etc. Head, hands and body can be localised in space using trackers, video techniques, gloves or suits. Trackers are devices that allow one to measure the position in space and the orientation of a small object, which is typically affixed to the body part that is to be followed. Video techniques use image recognition in order to follow, for example, the hand or body and reconstruct its position, orientation and posture from 2-D video images. Gloves and suits allow one to measure relative positions and angles of components of the hands or body, thus providing for the recognition of postures.

A very comprehensive directory of manufacturers of input technologies, of which a large part is devoted to gesture capture devices, is available in [2.4-3].

Among the evaluation criteria to be taken into account for the evaluation of gesture capture devices, besides cost and dependability, are the following ones:

- accuracy, which is the expected measurement error;
- range, which defines an area or volume in which measurements can be done (accuracy is often specified for a given range);
- precision, i.e., measurement repeatability;
- resolution, which quantifies the smallest measurable physical change;
- update rate, which is the measurement frequency;
- and latency, that is, the time the system takes to report a physical change.

2.4.2.1.2 Contact Devices

These devices work in two dimensions and are mostly used through a straightforward translation into a two-dimensional space for pointing co-ordinates or selecting items in menus. These are classic interaction modalities and we will not deal with them much. Such devices can however be used for more advanced purposes, such as two-dimensional gesture recognition or handwriting recognition.

Some of these devices, e.g. graphic tablets, allow one to use contact to determine the beginning and end of gestures, which is in general a difficult problem.

But the main problem is that gesture is quite constrained. In particular, one hand is generally completely involved.

Two kinds of pointing devices can be distinguished: direct ones, which allow one to point on the screen surface, and indirect ones, for which interaction is mediated by a translation into the screen space. Indirect pointing devices require additional co-ordination in that the operator has to match his movements with displacements in a different plane.

Direct pointing devices include the following:

- Lightpens are attractive but must be picked up, lead to arm fatigue (if the screen is vertical), and obstruction of the screen by the hand. They are rather fragile devices.
- Touchscreens (capacitive, ultrasonic, resistive or using a matrix of light beams) are fairly easy to use and robust. Although they allow good precision, since the fingertip is very sensitive and accurate, it is nearly impossible to be precise on first contact (correction is needed). One way to achieve this aim would be to have the screen "sense" the finger as it approaches it, provide adequate (e.g. visual) feedback and perform an action only at contact; this has been tested on prototypes but is not an available technology yet. Other prototype touchscreens allow actions involving more than one finger and even sense forces tangential to the screen surface but have not been successfully developed either. The same arm fatigue and screen obstruction problems as with lightpens occur, with screen smudging in addition.
- Styluses (used in some notebooks) are more comfortable and precise. They allow comfortable handwriting but must be picked up and the arm fatigue problems would occur if they were not typically used on small screens where the hand has a resting point.

Among indirect pointing devices are the following ones:

- The mouse (optical, physical, or acoustic) is very precise and rapid but must be grasped and needs some available horizontal space; movement can be hampered by the wire, except for modern infrared-equipped models.
- Trackballs have the same use as mice but occupy less horizontal space. They are fast but need more training than mice do.

- Joysticks are fast and efficient for direction changes and small movements: they are thus good for tracking targets. Some force feedback is possible. *Absolute* joysticks map the position of the cursor, while *isometric* or *velocity-controlled* joysticks map pressure on the stick to velocity of the pointer. An example of the latter is the finger-operated mouse replacement found on some portable computers.
- Graphics tablets (resistive, magnetic, or acoustic) offer good performance for writing or drawing; modern models are sensitive to stylus pressure, allowing for very elaborate expression possibilities. These devices are comfortable and precise but use a lot of horizontal space.
- Touchpads present the same advantages as touchscreens, without obscuring the screen. Some training is required to reach a correct co-ordination.

A more detailed summary of contact devices can be found in [2.4-4, pp 59-63] and [2.4-5, pp 244-253].

2.4.2.1.3 Trackers

Overview. We will present here an overview of various devices allowing one to measure in real time the position of an object in space — that is, the six parameters (three co-ordinates and three angles) that correspond to its six degrees of freedom. These devices can be used for head localization or hand tracking (possibly in conjunction with a glove). They have also been used for person localization and body posture recognition, although in the last case the number of devices affixed to the body can make such systems awkward.

Tracking can be done using either mechanical binding to potentiometers, or non-contact techniques such as magnetic fields, ultrasonic or infrared beams, or radars.

Mechanical Tracking. Mechanical tracking involves permanent mechanical binding of the followed object to its environment, using potentiometers linked to the object either by articulated rods or taut cables. This allows very fast refreshment rates (up to 300 Hz) and very small latency. It is also an inexpensive solution. On the other hand, usable range is small and the apparatus impairs free movement; the bulk and attachments to the body preclude any in-flight use and make these systems difficult to accept. This type of tracker has mostly been used for head localisation.

Electromagnetic Tracking. Electromagnetic trackers include an emitter, which is made of three coils radiating alternating electromagnetic fields in a radius of a few meters. The receiver, which is the mobile element, is also made up of three coils, which receive signals varying with its position relative to the emitter. An electronic unit ensures proper modulation of the radiated fields, measurement of the currents in the receiver coils, possibly filtering of these data, and computing of the position. Early models (ca. 1987) used an analog technology, which led to a long response time and a very noisy output; that is, even when the device was still, the output indicated movement. The introduction of oversampling and digital signal processing provided better results. Some of these devices allow simultaneous measurement of the position of several receiver units.

These trackers are costly, but on the whole they are the most precise among the no-contact ones. They also offer quite a large operating range.

The main disadvantage of electromagnetic trackers is that any metallic object in the vicinity will generate induced electromagnetic fields, which will hamper measurements. Obviously, any source of electromagnetic fields such as a video monitor will introduce errors as well. Also, there must be an electric connection between the receiver and the electronic unit, which can limit free movement.

Ultrasonic Tracking. Acoustic trackers make use of ultrasonic pulses to compute distances from time propagation measurements. As for electromagnetic trackers, emitters and receivers come in groups of three; measuring all nine distances between emitters and receivers allows one to compute the position and orientation of the mobile element.

The main advantage of these trackers is that they work seamlessly in metallic environments. They also tend to be much less expensive than electromagnetic ones.

Contrary to the electromagnetic trackers, ultrasonic trackers face a directivity problem — receiver units have to "see" the emitter. The time delay is greater than with other trackers since it includes propagation of ultrasonic waves. Furthermore, since the speed of sound varies with temperature, temperature variations lead to errors. Other limits stem from the compromise that must be made in the frequency choice. Too high a frequency will decrease the range since air attenuates the ultrasonic waves (useful range at 80 kHz is limited to about 2 meters); the usable range will further be decreased since directivity increases with frequency. On the other hand, with low frequencies, precision is limited by wavelength (4 mm at 80 kHz). Some new trackers continually measure phase shift between source and reception, which leads to improvements in precision and delay.

Other problems with this tracking technology are its sensitivity to ambient sound perturbations, as well as to reflections of the ultrasonic waves off walls.

Optical Tracking. Optical trackers generally use infrared LEDs. Most of them are built for specific needs. They can be divided into those that measure angles using point receivers, e.g. phototransistors, and the ones that make use of planar receivers, mostly video cameras.

Planar-receiver-based devices measure the angular position of a point light source (or a reflective marker) using two cameras. Using markers relieves the need for an attached wire to provide electric feed, but makes image processing harder unless external illumination is provided. The cameras can either be fixed and track mobile markers (*outside in*), or be the mobile parts and thus look at fixed beacons (*inside out*). The *outside in* principle limits precision (the cameras must have a wide field of view, yet measure small movements); with the *inside out* principle, the results are better provided there are enough beacons in the environment, but the cameras' weight can be a problem.

These devices face the same directivity problem as ultrasonic trackers, and the usable range is similarly limited. Furthermore, they can be perturbed by light and the use of infrared light makes them impossible to use in

Table 2.4-I: Comparison of trackers

Type of tracker	Range	Precision	Cost	Comments
Mechanical	limited	very good	low	bulky
Electromagnetic	large	good	high	sensitive to magnetic fields and metal objects
Ultrasonic	visible area	low	low	sensitive to temperature, humidity and sound
Optical	visible area	adequate	variable	sensitive to light
Inclinometers, compasses	unlimited	adequate	low	no position measurement
Gyroscopes, accelerometers	unlimited	low (integration)	low	shock-sensitive

combination with night-vision goggles. In a military context, possible remote detection of infrared sources can also be a cause for concern.

Other Trackers. More recently, some emitter-less trackers have begun to appear, using similar principles as plane and missile guidance systems, such as inclinometers, Hall-effect compasses, gyroscopes or accelerometers. Inclinometers measure orientation using gravity to determine the vertical direction (and are thus sensitive to acceleration). Compasses find the North magnetic pole using Hall effect; they are perturbed by magnetic fields and metallic masses. Gyroscopes either use rotating masses or piezo-electric crystals. They only allow relative measurement, as do accelerometers; this leads to accumulating integration errors when absolute position must be computed and hence frequent recalibration is needed.

Table 2.4.I gives a summary of different tracking technologies.

2.4.2.1.4 Video-based Systems

Video-based systems either work by tracking markers or by identifying silhouette features in images. They have been used for hand gesture recognition and body posture detection; they are the least intrusive method for the latter.

Marker-based Systems. These systems generalise the previously mentioned optical trackers; they typically use infrared LEDs or reflecting dots (which may require specific illumination) and work by correlating the marker positions detected by several cameras to compute tri-dimensional co-ordinates. Problems include the limitations on the number of markers due to the computational complexity of correlations and the fact that markers positioned too closely to one another cannot be discerned because of the limited resolution of the cameras.

Computational Vision Systems. These systems use classical image recognition techniques to find silhouettes of the hands or body and identify postures. The computational problems are even worse than with marker-based systems if real-time operation is required. Limited camera resolution leads to a compromise between adequate recognition of tiny elements (such as fingers) and large field of view which is necessary for free movement. Obstruction (e.g. of the fingers by one another or by the hand) is another problem; and correlating several sources in order to compute three-dimensional information, though a workable solution for simple gestures such as pointing [2.4-6], is far from trivial.

A common problem to both of these techniques is that even 60 frames per second (the current limit for typical video cameras) are not enough to follow rapid hand movements.

2.4.2.1.5 Gloves

Gloves measure hand and finger angles and movements of the fingers relatively to the hand. Most of them can be equipped with a magnetic position tracker in order to follow the global hand position. Numerous sensors are needed and lots of data are thus issued. Various technologies can be used, including optic fiber, Hall effect, resistance variation, or accelerometers. Gloves have been used as pointing devices, but they open a much richer field of interaction through hand posture recognition and even dynamic gesture symbolic interpretation. The main problems encountered are repeatability, precision and reliability. Almost every glove needs calibration before each use, since the way it is fitted onto the user's hand greatly affects the measurements.

Sensor technologies used in gloves have tentatively been applied to body posture recognition through so-called data suits, but this field is still a research domain.

The first widely known glove, the Dataglove, appeared on the market in 1987. It takes advantage of the attenuation of light in bent optic fibers to compute the joint flexions. It uses 10 sensors (two on the lower joints of each finger, two on the thumb) and works at 60 Hz. Its accuracy is on the order of 5 degrees; it is limited because attenuation is not a linear function of the angle. This precision is insufficient for complex gesture recognition. The main drawback of this technology is that light attenuation becomes permanent after some use and the fibers thus need replacement. They are also quite fragile. Production of this glove is discontinued.

A subsequent model, the Cyberglove (1990), uses 18 or 22 strain gauges. There are two for thumb joints, two or three for finger joints, four for abduction (thumb, middle-index, middle-ring and ring-pinkie), two for palm arch (thumb and pinkie) and two on the wrist (pitch and yaw). The operating rate can be a bit more than 100 Hz and the accuracy is about 1 degree. This glove is quite expensive but allows very good performance.

Game designer Nintendo introduced the Powerglove in 1989 as an intended game controller. This is a very inexpensive device that uses the variation of conductivity of carbon ink tracks to measure bend. It is coupled with a low-price ultrasonic tracker. Production has been stopped, not due to the modest performance, but because the dedicated game market was not developed enough.

Table 2.4-II: Comparison of glove technologies

Glove technology	Precision	Cost	Comments
optic fiber	low	high	fragile, subject to wear
strain gauges	high	high	
resistive ink	very low	very low	
Hall effect	high	very high	cumbersome
accelerometers	low for position measurement	research	sensitive to acceleration and shock

A considerably more elaborate model is the Dexterous Hand Master (1990), which tries to alleviate the slipping problems by reinforcing the bonds between hand and sensors. To this end, the DHM includes an exoskeleton in the joints of which Hall effect sensors (up to 4 per finger) measure joint angles with a frequency of up to 200 Hz. Sensitivity and resolution are high, but calibration problems remain. Also, the glove is heavy (350 g.) and it is not known whether the metallic exoskeleton is compatible with the presence of a magnetic tracker for hand position measurement.

The SensorGlove [2.4-7] is a more recent and experimental device, which uses accelerometers. It does not allow accurate position measurement (the double integration needed leads to accumulating errors), but is usable for dynamic gesture recognition. Accelerometers allow a high time resolution (up to 5 kHz) and are lightweight devices, but they are sensitive to shock. Also, it is not clear how they behave in high-acceleration environments such as a cockpit.

Table 2.4.II summarises essential aspects of different glove technologies.

2.4.2.1.6 Other Devices

3D mice are devices dedicated to moving a pointer in a three-dimensional space. They can use the same kind of technology as trackers (i.e. cables bound to potentiometers, magnetic tracking, ultrasonic or infrared beams) but are typically designed as generalisations of mice. Typical 3D mice allow one to move in about a 25 cm radius with a precision of less than 1 mm and a measurement frequency of 250 Hz.

Spaceballs, sometimes also known as trackballs, are spheres that allow 6 degrees of freedom (applied force on each axis and torque around each axis). These are rather inexpensive devices. The main problem is whether real independence between these 6 degrees can be attained by the user — it is hard to apply a linear force with no torque at all.

In order to widen the usable range of gesture recognition, "smart ceilings" [2.4-8] and "smart floors" are being investigated. Smart ceilings use a network of LEDs and head-mounted photo-receivers and allow to detect the position and orientation of the operator in a whole room. Smart floors consist of a matrix of floor-mounted pressure sensors and give information on the position of the user, but could also determine his movement direction and speed, and also be a hint for user identification since they allow one to estimate weight.

2.4.2.1.7 Feedback

A number of applications of gesture, particularly in the fields of virtual reality and computer-aided design, have shown the importance of feedback in gestural interaction. (we mean here only tactile and force feedback, as opposed to visual feedback, which is not always possible (e.g. in "eyes-out" applications) and can even be a nuisance by inducing perturbations [2.4-9]). As a matter of fact, identifying a virtual object, grasping it and manipulating it precisely are very difficult if no perceptive information is provided. Feedback information helps a lot in feeling in command of the system. For instance, it has been used as a guidance tool in window managers, using a magnetic mouse whose resistance can be programmed and thus signal for example when the mouse crosses a window limit.

But feedback can also change the interaction in more fundamental ways, such as in the following example from [2.4-10]: the operator draws figures on a map and his/her drawing tool receives a force feedback proportional to the population density gradient. This immediately allows e.g. the drawing of a possible least disrupting highways by simply following the paths of least resistance.

A distinction must be made between tactile and force feedback. The first one addresses the tactile sense and provides information on the nature of the surface of a grasped object (geometry, roughness, temperature) while the second one involves the proprioceptive sense and informs on the elasticity, weight and movement of the object.

Evaluation criteria include bandwidth, which determines, for example, the quality of a simulated texture, and power of the force feedback. There is a compromise to be reached here, since great forces are needed to simulate a hard object, but misapplied or too strong forces could be harmful to the user. Delay is also quite an important factor: force feedback arriving too late is useless and can even make a system unusable.

Determination of adequate stimuli to be sent to the user is as yet rather primitive; it is mostly either based on recording of real forces during the manipulation of actual objects, which is an ad hoc technique, or on rough estimations of the stimuli as functions of distances to the objects.

Tactile Feedback. Pneumatic, shape-memory and vibro-tactile technologies have been used for providing tactile feedback. Experiments have also been performed using hydraulic systems, electric stimulation of the skin or even

direct neuro-muscular stimulation. The currently available devices are few; this area is still mostly a research domain.

Pneumatic devices use a number of small balloons, generally integrated in a glove, and which can be inflated to apply pressure on the fingers or palm.

A matrix of micro-rods made of a shape-memory material has been used for tactile stimulation; the rods change shape when adequate heating is applied and are suitable for miniature devices.

Vibro-tactile devices use small loudspeakers, electromagnetic or piezo-electric micro-rods, which transmit audio-frequency (around 200 Hz) vibrations to the skin. They are most appropriate to simulate texture of a virtual object. Some experiments have added thermal stimulation to these, e.g. in order to indicate emergency conditions.

Finally, handles can be equipped with tactile feedback systems; applying small modulated forces to the handle allows one to simulate a texture or viscosity.

Force Feedback. Force feedback systems can use electric, hydraulic or pneumatic technologies. They have first been applied to telemanipulation arms; since then, increasing miniaturisation has allowed one to fit such systems into sticks and even gloves and joysticks. Yet the main disadvantage is that most of these systems remain bulky and rather intrusive, which prevents their use in transportable or wearable devices.

2.4.2.1.8 Software Techniques

Algorithms. A variety of algorithms have been used for the interpretation of gesture data.

Posture recognition typically uses ranges of values as criteria, classifying the posture among one of those known if all data points measured by sensors fit in appropriate value intervals deduced from training gestures. More sophisticated techniques involve matching vectors of values with some (e.g. Euclidean) distance as a criterion. Hidden Markov models and neural networks have also been used.

Dynamic gesture recognition is a much harder problem and has also been addressed using hidden Markov models and neural networks, as well as dynamic time warping for addressing the problem of matching gestures to an internal model. Rubine [2.4-11] has designed a dedicated algorithm for geometric feature extraction that presents very good performance and efficiency.

A very hard problem is segmentation: as is the case with continuous speech recognition, co-articulated gestures interfere with each other and thus with the detection of individual gestures; it is also a nontrivial problem to identify the beginning and end points of a gesture. Typical solutions require the operator to take a "default" hand posture between gestures, which serves as an anchor for the system. Davis and Shah demonstrate the feasibility of using simple finite state machines under this paradigm in [2.4-12].

Interfaces. Numerous experiments have shown that using gesture for interacting with systems calls for new interface paradigms, of which we present two examples drawn from ongoing research.

Cadoz presents in [2.4-1] what he calls *instrumental communication* with the computer. The concept of instrumental gesture is derived from the use of gesture in musical practice. It is a combination of the ergotic and semiotic gesture, in that it is as well a physical manipulation of an object as a means of expression through the dynamic control of various physical processes. The problem is then, given a task to be controlled, to define a communication instrument whose operation is able to support the communication needs of the task. An example is given in section 2.4.3.

Marking menus [2.4-13] are a sort of combination between stroke recognition systems and pie menus. The user can draw a stroke corresponding to a command, but if s/he stops drawing, a menu will appear in order to show possible commands and the corresponding strokes. Novice users can thus interact with the system easily and get a reminder of the fast, equivalent command every time they select it from the menu, which gradually leads them to a more efficient interaction. This paradigm has proven an efficient way of learning abstract two-dimensional commands.

2.4.2.2 Facial Gesture

A human face provides a variety of different communicative functions, such as identification, perception of emotional expressions, and lip-reading. Identification is not addressed in this document. Lip-reading is discussed in the speech state-of-the-art, section 2.1.3.2. Face perception is currently an active research area in the computer vision community. Much research has been directed towards feature recognition in human faces. Three basic techniques are commonly used for dealing with feature variations: correlation techniques, deformable patterns, and spatial image invariants. Several systems for locating faces have been reported. By moving a window covering a subimage over the entire image, faces can be located within the image. Sung and Poggio [2.4-14] report a face detection system based on clustering techniques. The system passes a small window over all portions of the image, and determines whether a face exists in each window. A similar system with better results has been reported by Rowley *et al.* [2.4-15]. A different approach for locating and tracking faces is described in Hunke and Waibel [2.4-16]. This system locates faces by searching for skin color. After locating the face the system extracts additional features to match this particular face. Eigenfaces, obtained by performing a principal components analysis on a set a faces is commonly used to identify faces.

A by-product of face recognition research provided a valuable and novel communications interface for the non-vocal handicapped [2.4-17]. The system was adapted to recognise three expressions: mouth open; mouth closed; and mouth open with tongue extended. The motivation was to provide a computer interface for a child with cerebral palsy. The child had no muscle control from the neck down, which also rendered her incapable of speaking. From the neck up, the child has only limited control. Although she cannot speak, she can manage facial expressions. She indicates "yes" by opening her mouth, "no" by extending her tongue, and "null" by closing her mouth. Laboratory results with normal subjects indicate that these symbols can be distinguished with a greater than 95% accuracy.

Another active research and development area is the recognition of facial expressions. This work combines techniques for tracking and locating the face with the recognition of different expressions such as disgust, anger, happiness, and surprise. The goal is to develop an intelligent interface that would adapt to the user based on the emotional state determined from the his/her facial expression. Examples of this work is that of Essa and Pentland [2.4-18] and Yacoob and Davis [2.4-19].

2.4.3 APPLICATIONS TO DATE

2.4.3.1 Command and Control

A typical example of innovative application of a classical gesture interaction device, the mouse, was a Macintosh add-on that could recognise some "doodles", strokes drawn on the screen with the mouse and accordingly launch a command. The system could be configured for a preferred set of commands and trained by the user.

More recently, in [2.4-9], use of a touchscreen as a generic command means for various car equipments has been proposed in a prototype car. The system consists of a touchpad where symbols can be drawn with the fingertip in order to launch specific commands, e.g. radio selection, navigation help or for phone dialling. It is designed to be used without any visual feedback and experiments show that it is quite possible to draw symbols or letters on such a surface without seeing the result. As a matter of fact, the quality of hand-written letters and symbols is demonstrably lower when there is a visual feedback: the operator has then a tendency to over-compensate his/her drawing, which leads to trembling and ill-drawn shapes.

2.4.3.2 Remote Manipulation

Remote manipulation of objects (be they too heavy or dangerous, e.g. radio-active), has been done either using direct mechanical transmission (pantographs) or electric engines, which allows amplification. Even if it does not actually include a computing system, it involves the transmission of gestural information; in that respect, it is a forerunner of a number of applications.

For instance, the GROPE project of the University of North Carolina, presented in [2.4-20] and [2.4-21], is among the first applications to have used force feedback not for remote manipulation, but for interacting with a computer program. The studied domain is the simulation and graphical representation of interactions between complex molecules. A force feedback manipulating rod allowing six degrees of freedom has been specifically developed; when the user modifies the simulated position of one molecule by moving the rod, the simulation computes inter-molecular forces and sends them back through the force feedback system. Despite relatively limited performance (due in particular to the computational time needed for simulation), and thus not very realistic sensations, this system allows one to find out possible chemical bonds between molecules.

Quite a successful and original application was developed in Medialab [2.4-22], which consists of mediated animation of virtual characters. Facial motion capture devices allow one to map the operator's feature movements onto features of an animated character and thus to produce image sequences in a very short time compared with classical animation techniques.

2.4.3.3 Virtual and Augmented Reality

Virtual reality applications, which are investigated, e.g. in [2.4-23] and [2.4-24], require gesture interaction:

- for the purpose of interaction with the objects of the virtual world;
- in order to provide for the immersion feeling by updating in real time the visual representation of the virtual environment as a function of the users' movements, especially of the head movements. The so-called "immersion syndrome", in which every action of the user is considered intended for the system, is a feature of the system.

If the user is required to "touch" virtual objects in any way, then the accuracy of depth perception becomes an issue, particularly for computer generated displays which lack the rich variety of textural cues available to us in real life. Takemura, Tomono and Kobayashi [2.4-25], using a stereoscopic projector to display targets in three-dimensional space, found that subjects could "touch" the objects with a three-dimensional tracker with satisfactory accuracy. Ineson and Parker [2.4-26], however, using a similar task but with head-mounted display, found that while some subjects could "touch" the virtual objects with good accuracy, others had great difficulty in judging their depth.

Application fields are numerous; the most popular include simulation (especially in the military and medical fields), data visualisation, education and training, communication in a virtual environment and of course entertainment. Problems in this field include psychological immersion problems and the impossibility of direct interaction with the real world.

Augmented reality is a more pragmatic approach whose principle consists in adding data processing abilities to the familiar objects we interact with in the real world. It thus aims at integrating computing systems into the real world instead of embedding the user in a simulated world. An example is the Digital Desk [2.4-27], which allows one to work with classic paper but also to use at the same time digital tools. For instance, if the user is drawing, a video camera can scan the drawing, which can then be digitally edited by means of a projector; it is even possible to mix both and work on a hybrid document, partly real and partly electronic.

Some projects of virtual cockpit also work in this fashion by projecting synthetic imagery onto the physical environment of the pilot, making it possible to keep using the conventional controls while adding the power of interaction with the projected virtual tools. For example, White *et al.* [2.4-28] were interested in the problem of interacting with real cockpit instruments when direct vision of the instruments was obscured. They set up a virtual keypad on an HMD which overlaid a real keypad, and operated it using a finger-tracker to give feedback of the finger position to the user.

2.4.3.4 Sign Language Understanding

Human sign language understanding is largely considered a point of reference in the field of gesture recognition, due to the complexity and expressiveness of these languages. A large amount of current research is dedicated to human sign

language understanding, using pattern matching, neural nets, dynamic programming or Markov models. Since human sign languages as such are not adequate as control modalities, we will not detail this field of application in this document. Several papers on the subject appear in [2.4-29].

2.4.3.5 Various Applications

[2.4-1] presents the "Retroactive modular keyboard" developed at ACROE. It is a keyboard equipped with accurate position sensors and elaborate tactile and force feedback on each key, which allow for very realistic feedback. It is modular and can be configured to accommodate a given number of degrees of freedom. Coupled with a real-time complex physical object modelling system and a gesture editor, the system can simulate a variety of objects and actuators and various gesture vocabularies can be defined and recognised. Demonstration applications include vehicle control, musical operation and a robotic system.

CHARADE [2.4-30] is a system designed for computer-aided presentation using hand gestures. It allows the presenter to control a Hypercard presentation by means of hand gestures. Wearing a Dataglove, the speaker points at the screen, which constitutes an "active zone", and makes a short hand gesture corresponding to the required command; the gesture is then matched to an internal model consisting of a start position, hand and arm movement, and a stop position. This scheme prevents the immersion syndrome in that the speaker can keep using gesture in addressing the audience. It also alleviates the problems of gesture co-articulations and a careful choice of start and end positions makes recognition easier. The choice of tense postures as start positions and relaxed ones as end positions is also a helpful hint for the recognition device. A variant of the Rubine algorithm [2.4-11] is used. Sixteen commands such as "next/previous page", "next/previous chapter", "table of contents", "mark this page", or "highlight area" are available. An ad hoc, iconic notation was designed in order to write documentation for the system. Recognition scores of 90 to 98% have been reached by trained users.

2.4.3.6 Cockpit Applications

The choice of a suitable selection mechanism for a task is of great importance. The following two examples involve three-dimensional positional tracking, but with very different results.

Ineson, Parker and Evans [2.4-31] compared a video-based finger tracker with several other designation mechanisms to select buttons on a virtual, head-down panel during a low level flight simulation. Feedback for contact with the button was by colour change, and activation of the button required depressing a HOTAS switch. The finger-tracker was not liked since it removed the hand from the flying controls for a substantial length of time, and some subjects found the device awkward to use since it was necessary to keep the finger in clear view of the tracking cameras. Although the normal way of operating a button is to reach out and press it, the task is essentially two-dimensional. Pointing methods such as head designation and selection using the stick-top cursor controller are therefore suitable selection mechanisms, and both were in fact preferred to the finger tracker. Finger pointing direction would have been more suitable than finger position, since it could have been operated with the hand on or near the controls. Voice command was however the overwhelmingly preferred selection technique for this task.

A series of experiments carried out at WPAFB [2.4-32 - 2.4-35] required a true three-dimensional selection of targets from a head-down, three-dimensional tactical map. In this case an electromagnetic tracker was strapped to the back of the hand, resulting in a more robust and responsive tracking than the video-based technique used by Ineson *et al.* The tracking volume was remote from the actual map so that hand movements were actually made in a space close to the aircraft controls rather than within the volume of the map, and this hand volume was reduced in scale so that hand movements were small compared with the size of the map. The volume was divided into four depth planes, so accurate depth control was not required. The hand tracker worked well, and was in general faster and more accurate than a three-dimensional joystick, although if the tracking volume was made too small selection accuracy was impaired. Voice selection was also used in some of the experiments [2.4-32, 2.4-35], but unless the targets were labelled it was difficult to define a suitable vocabulary, and the method was slow compared with hand movement.

Reising *et al.* [2.4-32] and Solz *et al.* [2.4-33] also investigated two methods of simplifying object designation - contact cueing (colour change when the cursor was within the target volume) and proximity cueing (automatic selection of the target nearest to the cursor). The latter was found to be particularly helpful.

Not only must the device be chosen to suit the task, but the environment in which the device is to be used is a factor which cannot be ignored. A positional hand tracker might be the preferred device in a relatively benign environment, but might become unusable under the G and vibration levels found in a fast jet or a helicopter. A three-dimensional joystick which would have the advantage of supporting the hand, but the space needed to integrate such a device must then be considered. The glove required by a gesture recogniser might be incompatible with safety equipment or might interfere with other tasks requiring finger-tip sensitivity. System lags which are tolerable in a controlled experiment might become impossible when the user has to attend to several tasks at once. Environmental and integration issues such as these might severely restrict the choice of control devices for a specific task.

2.4.4 REQUIRED ENHANCEMENTS AND PROGNOSIS

Currently, the main disadvantage of existing gesture-capturing devices is they limit the users' freedom of movement. This can be due to the need for grasping a mobile part (contact devices, mechanical tracking) or limited sensors range (video cameras, magnetic and, even more, ultrasonic and optical devices). Miniaturisation progress in the fields of energy sources and electronic devices can help solve this problem.

Static posture recognition has made great progress and allows reasonably high recognition rates, provided the user performs a standard procedure such as pointing at a signal area or assuming a standard posture prior to issuing a command. This is not yet true for dynamic gesture

recognition and software techniques are still developing in this field. The main difficulty is detecting gesture's beginning and end points. Hints such as hand speed and tension are currently being investigated.

General interface problems such as immersion are still not solved in a general way. The definition of an active zone (in which gesture recognition operates) partly solves this problem but may not be adequate for all applications.

The development of adequate interface paradigms for gesture interaction with computers is still under active research; a consensus on the integration of gestures in interfaces is far from being reached.

In the domain of feedback, the determination of appropriate stimuli is also still largely in its infancy. Appropriate modelling of physical objects and of their interaction with body parts is a prerequisite.

Finally, to this day, learning how to operate a gesture-based interface is mostly done by example. Gesture notation is undergoing a lot of research. An example is HamNoSys (Hamburg Notation System) [2.4-36], which is a general iconic system; although it initially aimed for notation of human sign languages, it has since proved helpful in the design of artificial gesture languages.

2.4.5 REFERENCES

2.4-1 Cadoz, C., "Le geste canal de communication homme/machine, la communication 'instrumentale'", Technique et science informatiques, 13(1), 1994, pp 31-61.

2.4-2 So, R. H. Y., "Comparison of the transmission of vertical seat vibration to the head and finger in a stationary target aiming task", United Kingdom and French Joint Meeting on Human Response to Vibration, 1988.

2.4-3 Buxton, B., "A directory of sources for input technologies", 1998, available on the Web as http://www.dgp.utoronto.ca/people/BillBuxton/InputSources.html

2.4-4 Dix, A., Finlay, J., Abowd, G., and Beale, R., "Human-Computer Interaction", UK, Prentice-Hall, 1993 (ISBN 0-13-437211-5).

2.4-5 Shneiderman, B., "Designing the User Interface — Strategies for Effective Human-Computer Interaction", 2nd edition, Addison-Wesley, 1992 (ISBN 0-201-57286-9).

2.4-6 Fukumoto, M., Mase, K., and Suenaga, Y., "Real-time detection of pointing actions for a glove-free interface", in IAPR Workshop on Machine Vision Applications, December 7-9, 1992, pp 473-476.

2.4-7 Hofmann, F. G., and Hommer, G., "Analyzing Human Gestural Motions using Acceleration Sensors", in "Progress in Gestural Interaction", Proceedings of Gesture WorkshopÊ'96, University of York, March 1996.

2.4-8 Ward, M., Azuma, R., Bennett, R., Gottschalk, S., and Fuchs, H., "A demonstrated optical tracker with scalable work area for head-mounted display systems", in "1992 Symposium on interactive 3D graphics", Association for Computing Machinery, Cambridge, 1992, pp 43-52.

2.4-9 Kamp, J.-F., and Poirier, F., "Un dispositif tactile pour la commande en véhicule : étude d'une utilisation sans retour visuel", in "9èmes journées sur l'ingénierie de l'interaction Homme-Machine (IHM'97)", 1997 (ISBN 2-85428-459-3).

2.4-10 Negroponte, N., "Being digital", Coronet, Hodder & Stoughton, 1995 (ISBN 0-340-64930-5)

2.4-11 Rubine, D., "The automatic recognition of gestures", Ph. D. thesis, Carnegie-Mellon University, 1991.

2.4-12 Davis, J., and Shah, M., "Gesture recognition", University of Central Florida Technical Report CS-TR-93-11, 1993.

2.4-13 Kurtenbach, G., and Buxton, W., "The Limits of Expert Performance Using Hierarchic Marking Menus", in ACM and IFIP joint conference on Human Factors in Computing Systems (INTERCHI'93), pp 482-487, 1993.

2.4-14 Sung, K. and Poggio, T., "Example-based learning for view-based human face detection", Technical Report 1521, MIT AL Lab., 1994.

2.4-15 Rowley, H. A., Baluya, S., and Kanade, T., "Human face detection in visual scenes", Technical Report CMU-CS-95-158, CS Department, CMU, 1995.

2.4-16 Hunke, M. and Waibel, A., "Face locating and tracking for human computer interaction", in 28th Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 1994.

2.4-17 Goble, J. R., Suarez, P. F., Rogers, S., K., Ruck, D. W., Arndt, C., and Kabrisky, M., "A facial feature communications interface for the non-verbal", IEEE Trans. On Engineering in Medicine and Biology, Sept 1993.

2.4-18 Essa, I. A. and Pentland, A. P., "Coding, analysis, and recognition of facial expressions", Technical Report No. 325, MIT Media Lab, April 1995.

2.4-19 Yacoob, Y. and Davis, L. S., "Recognizing Facial Expression", Technical Report CS-TR-3265, University of Maryland, Computer Vision Laboratory, May 1994.

2.4-20 Brooks, F. P., "Grasping reality through illusion: Interactive graphics serving science". In "Human Factors in Computing Systems", Proceedings of CHI '88 -, ACM, New York, 1988, pp 1-11.

2.4-21 Brooks, F. P., Ouh-Young, M., Batter, J. J., and Kilpatrick, J., "Project GROPE: Haptic displays for scientific visualisation", Computer Graphics, 24, 4, August 1990.

2.4-22 Limantour, P., "Medialab: Masters of Motion Capture", Computer Graphics World, October 1996.

2.4-23 Krueger, M., "Artificial Reality", 2nd edition, Addison-Wesley, 1991.

2.4-24 Burdea, G., and Coiffet, P., "La réalité virtuelle", France, Hermès, 1993 (ISBN 2-86601-386-7).

2.4-25 Takemura, H., Tomono, A., and Kobayashi, Y., "A study of human-computer interaction via stereoscopic display", in "Work with Computers: Organisational, Management, Stress and Health Aspects", pp 496-503, M. J. Smith and G. Salvendy (eds), Elsevier Science Publishers, 1989.

2.4-26 Ineson, J., and Parker, C. C., "The accuracy of virtual touch", DRA Working Paper DRA/AS/MMI/WP95036/1, 1995.

2.4-27 Wellner, P., "Interacting with Paper on the DigitalDesk", Communications of the Association for Computing Machinery, 36, 7, July 1993, pp 87-96.

2.4-28 White, J. L., *et al.* "Virtual cockpit concepts: an evaluation of data entry techniques", DRA Report DRA/AS/MMI/CR95168, 1995.

2.4-29 "Progress in Gestural Interaction", Proceedings of Gesture WorkshopÊ'96, University of York, March 1996.

2.4-30 Baudel, T., and Beaudoin-Lafon, M., "Charade: Remote control of objects using free-hand gestures", Communications of the Association for Computing Machinery, pp 28-35, July 1993.

2.4-31 Ineson, J., Parker, C. C., and Evans, A., "A comparison of head-out and head-in selection mechanisms during simulated flight", DERA Customer Report DERA/AS/SID/510/CR97153, 1997.

2.4-32 Reising, J. M., Liggett, K. K., and Hartsock, D. C., "Exploring techniques for target designation using 3-D stereo map displays", International Journal of Aviation Psychology, 3(3), pp 169-187.

2.4-33 Solz, T. J., et al., "3-D stereo displays: how to move in the third dimension", SPIE conference, Orlando, May 1995.

2.4-34 Solz, T. J., et al., "The use of aiding techniques and varying depth volumes to designate targets in 3-D space", Proc. 38th Human Factors Soc. AGM, 1994.

2.4-35 Reising, J. M., et al., "New cockpit technology: unique opportunities for the pilot", SPIE conference, Orlando, April 1994.

2.4-36 Pritzwill, S., Leven, R., Zienert, H., Hanke, T., Henning, J., *et al.*, "HamNoSys (version 2.0) Ñ Hamburg Notation System for Sign Languages / An introductory guide", in International Studies on Sign Language and the Communication of the Deaf, vol. 5, Hamburg, Germany, 1989.

2.4-27 Sturman, D. J., and Zeltzer, D., "A survey of glove-based input", IEEE Computer graphics and applications, 14, 1, January 1994, pp 30-39.

2.5 BIOPOTENTIAL-BASED CONTROL

2.5.1 INTENTION OF THE TECHNOLOGY

Biopotentials can be measured from the natural electrochemical activity of many physiological systems. These signals are produced when excitable cells, such as muscle or nerve cells, are stimulated in response to an internal or external stimulus.

We are interested in two types of biopotentials, the electromyographic (EMG) signals associated with the contraction of skeletal muscle and the electroencephalographic (EEG) signals associated with brain activity, because these signals can be modified voluntarily to indicate the intention of the operator. As a control modality, the principal objective is to measure biopotential activity from the operator so that it can designate desired control options or augment other control modalities to reduce selection ambiguities.

Processed EMG and EEG signals can also be used to monitor operator alertness, muscle fatigue and workload. Because EMG signals are proportional to muscular effort and do not require movement of the limb, they can be measured in situations, such as high-g environments, where limb motion is restricted. As well, both EMG and EEG signals precede actual physical movements and provide information on human sensory processing and on the preparation for motor activity (movement).

2.5.1.1 Why Appropriate as a Human-Machine Control Modality

In addition to applications as an assistive technology for persons with physical disabilities, biopotential-based control has a variety of potential applications in aerospace environments. These environments fall into two broad classes: (1) ones in which there are constraints on control access and (2) ones in which there are high manual workload demands. An environment which requires an operator to wear protective gear against chemical and biological agents is an example of the first class. The bulky clothing and gloves make it difficult, if not impossible, to operate small switches and controls. Extravehicular operation in space is another example. In addition to the limitations of the space suit, operators are constrained by the need to control the acceleration of their body when using large tools and controls. A third example is high acceleration flight in which g-forces essentially limit pilot access to all controls except the joystick and throttle. The Hands-On Throttle and Stick (HOTAS) system is, in part, a response to the movement limitations of high acceleration flight. While it is reasonably effective in this regard, it has created a new set of training and memory requirements that are discussed in Chapter 1.

A variety of aerospace applications fall into the second class, ones in which there are high manual workload demands. Maintenance technicians must devote high visual and manual attention to the task at hand. Frequent access to technical reference material is also required. Head-mounted displays and wearable computers are being developed to provide this information (see Section 4.3). However, the technician needs some means to interact with the information system, while keeping their hands devoted to their work. Voice control provides one option, but it can be constrained by high noise, the requirement for concurrent communication, as well as a variety of speaker characteristics (see Section 2.1). Biopotential-based control provides another option for such systems.

The control of secondary systems in flight can generate a high manual workload for the pilot. Although it is difficult to imagine a day when pilots might use biopotentials as a primary control, it is easier to foresee the use of biopotentials as a secondary control by which pilots or navigators perform multifunction display operation, weapons selection, radio frequency switching, or target selection.

Operator state monitoring represents a third class of potential applications, but one that does not involve explicit system control. In this case the biopotentials provide on-line data, not otherwise available, about the operator's physical and cognitive state. Research in this area has emphasised three domains of human-system interaction that are of strategic importance to aerospace operations: operator workload monitoring, error prediction, and physical and cognitive fatigue monitoring. In addition to passive operator state monitoring, several advanced interface programs have considered the use of operator state data as part of an interface adaptation scheme. If used in this manner, biopotentials would provide an implicit system control function that blurs the distinction between monitoring and control.

2.5.1.2 Human Capabilities and Limitations with Respect to this Control Modality

It has been a goal of control system designers in many fields to tap our natural physiological systems to achieve intuitive, non-fatiguing control of external devices. The idea of an operator using natural motions of their hand to teleoperate a dextrous robotic hand is one example where this intuitive mapping could reduce operator training and workload. Similarly, the notion of operating a device simply by thinking about the desired action represents the ultimate in intuitive control. Although current technology limits our ability to achieve such natural control systems, many practical devices have been designed and other promising technologies are being evaluated in the research community. The most widely studied biological signals for this type of control application are the EMG and the EEG. While the biological processes and recording methods are similar for EMG and EEG signals, each has distinct advantages and limitations when used as a control input. These are discussed separately, below.

2.5.1.2.1 EMG-based Control

As a muscle fibre contracts, there is an associated ionic exchange across the fibre membrane. The volume conduction of the resulting electric field from many active muscle fibres may be measured, via electrodes, as myoelectrical activity. We have volitional control over the level of myoelectric activity from normal skeletal muscles such as those in the arms and face. It is this voluntary control of the signal level which is the basis for EMG-based control of artificial limbs and orthotic devices for the disabled [2.5-1]. A convenient means of observing myoelectric activity is by an EMG

recording on the surface of the skin. The recorded signal is referred to as either the EMG signal or the myoelectric signal (MES).

Surface-recorded EMG signals occupy the 20-500 Hz frequency band and are in the hundreds of microvolts to tens of millivolts amplitude range. Both of these characteristics present practical recording problems. The peak of EMG signal power is close to the power line frequency and the EMG signal amplitude is far less than the electrical interference due to capacitive coupling between the body and power mains.

Biological instrumentation amplifiers use a differential arrangement to reduce the power line interference, but the interference rejection is only effective when the electrodes and skin are in contact and when the electrical impedance of the skin is low. Another problem associated with surface electrodes is that if they move relative to the skin surface, a noise signal is produced which can be confused with the true biological signals. In severe cases, this motion artefact completely overwhelms the EMG signal and looks to the control system as a large control signal. To minimise this unwanted effect a good electrode/skin contact must be maintained. Many current EMG-based control systems use small stainless steel electrodes with integrated biological signal amplifiers. These dry electrodes rely on continuous contact with the skin to produce a layer of perspiration to reduce the impedance of the electrode/skin interface. In a myoelectric prosthesis this is achieved by tailoring each socket to fit the residual limb of the amputee. For aerospace applications a method of integrating the electrodes into the helmet or clothing to ensure intimate electrode contact will need to be developed.

The EMG signal can be modelled as band-limited random noise with the amplitude modulated by the level or effort of the muscle contraction (Figure 2.5-1). Commercial EMG-based control systems use simple signal processing operations (rectification and low pass filtering) to extract the amplitude information from this signal. There is a trade-off between system response time and control accuracy. A long filter time constant will result in a sluggish system while a short time constant will result in a loss of control due to fluctuations in the signal. For a system with a reasonable response time, this fluctuation limits the number of functions which can be controlled by a single control channel to two. This is sufficient for the control of only a single degree-of-freedom (DOF). Even then the operator must be trained to reduce function selection errors. For example, an amputee will receive several hours of training before he/she becomes a proficient user of an artificial limb. Once this occurs, however, the limb is controlled in a very natural and intuitive manner.

Introducing more control channels by sampling the EMG signal from other muscles increases the number of DOFs which can be controlled but at the expense of extended user training. The difficulty arises because the user must follow a specific sequence to achieve control of a particular DOF. For example, in one popular system, [2.5-2] the amputee must co-contract biceps and triceps muscles to switch the control from an electric hand to an electric elbow. Once switched, the elbow is controlled by contracting the biceps to flex and triceps to extend. Control returns to the hand if the amputee relaxes both muscles for a specified length of time. Although

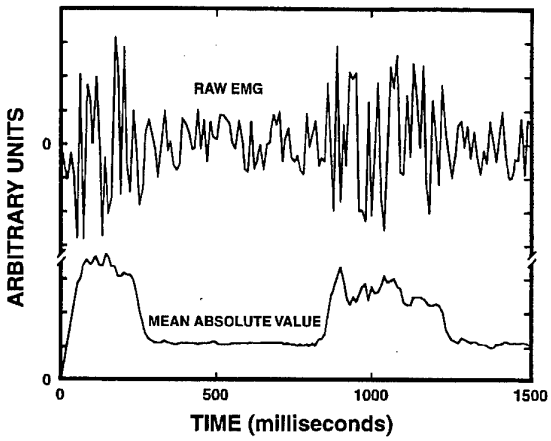


Figure 2.5-1 Raw and Processed EMG Signals Showing Two Brief Contractions.

many successful fittings have been achieved using this system, it (and similar systems) requires the user to learn to generate the specific signals required by the control system. User training is an issue in such systems because the natural way of activating a muscle involves a complex pattern in which many muscles are recruited to produce not only the desired movement, but to stabilise joints and provide resistance as well. To isolate individual muscles requires much concentration. The focus of current research is on systems which can learn to recognise the EMG signals generated by the user during these natural contraction patterns. These systems attempt to move the burden of control from the user to the control system. The assumption is that the user then controls the artificial arm in a more intuitive manner which improves operational accuracy and efficiency. For example, an amputee could choose to control an electric wrist by training the system to recognise the EMG signals which he generates when he rotates his residual limb. Likewise, they could train the same system to associate the EMG signals generated during other natural movements to control other prosthetic limb functions such as elbow flexion and extension and hand open and close. It is obvious that this type of control requires an intelligent control system with some means of retaining the user-specific control patterns. Pattern recognition systems under investigation [2.5-3] are based on conventional digital signal processing (DSP) chips and microprocessors such as those available from Texas Instruments and Motorola.

Even these systems, however, are limited by the low information content of the EMG signal and the limits set on the system by operator error. The best systems now achieve close to 100% control accuracy of, at most, four DOF [2.5-4]. A further limitation is in how these functions are controlled. Although simultaneous control of more than one DOF has been demonstrated, [2.5-5] most systems require the control of complex movements to be done in a sequential manner where each DOF is controlled separately. This increases the time required to do any task requiring the selection of more than one DOF.

2.5.1.2.2 EEG-based Control

In this chapter, we use the term EEG to include a variety of electrical signals recorded from the surface of the scalp. Although not formally correct, we will include transient and

steady-state evoked response approaches in the category of EEG-based control.

EEG recorded from the surface of the scalp represents a summation of the electrical activity of the brain (Figure 2.5-2). Although much of the EEG appears to be noise-like, it does contain specific rhythms and patterns that represent the synchronised activity of large groups of neurones. A large body of research has shown that these patterns are meaningful indicators of human sensory processing, cognitive activity and motor control. In addition, numerous EEG patterns can be brought under conscious voluntary control with appropriate training and feedback. The EEG signals of interest are in the 1-40 Hz frequency range with amplitudes ranging from 1-50 microvolts. Because of their small size, EEG signals are highly susceptible to contamination from eye and muscle activity, from external electrical sources and from movement of the user. These challenges can be managed, even in flight environments, but they require significant care and expertise on the part of system designers and operators.

Current prototype EEG-based control systems use scalp electrodes that were developed for clinical EEG recording. Electrical contact is maintained with a conductive paste. Aerospace applications require the development of convenient dry electrode systems, since it is unlikely that operational users will accept any significant electrode preparation and application process. Current prototypes being developed in Israel [2.5-6], the UK [2.5-7] and in the USA [2.5-8] suggest that this issue may be solved in the next 2-3 years, but a convenient, reliable and inexpensive system is not yet commercially available.

EEG-based control must address many of the same challenges as EMG-based controllers. As discussed above, designers face a significant trade-off between achieving short response times and smooth control outputs. Several factors contribute to this problem. First, the specific EEG pattern being used for control is typically a small component of the overall EEG signal and must be discriminated from normal background EEG activity. This signal processing takes time and can degrade system response. Second, various signal filtering schemes are commonly used to eliminate the sources of artefact listed above. These filtering steps smooth the control output, but can also introduce lag or delay. Finally, approaches which require the user to voluntarily modulate or produce specific EEG-patterns have an additional source of variability that must be managed. While users are able to rapidly raise a specific EEG component above a set threshold, holding it in a stable state is difficult. The EEG output typically shows brief drops below threshold that must be managed by the control algorithm. The required signal averaging or smoothing adds additional lag or delay to the system. These limitations, while severe, are not unique to biopotential-based control. Most eye-gaze-based controllers face many of the same problems in discriminating intentional from spontaneous eye movements.

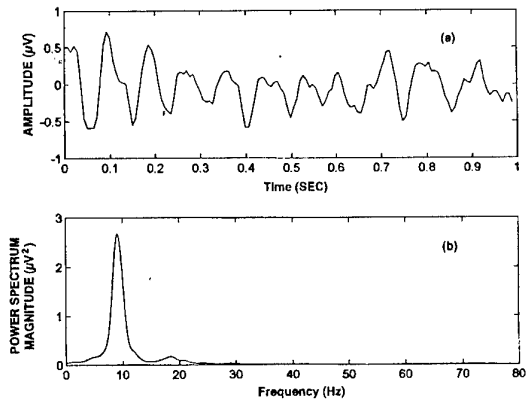


Figure 2.5-2 Sample Raw EEG Signal (a) and Power Spectral Density (b). Spectrum based upon 10 seconds of data from one subject with 1 second of raw EEG shown. High-pass filter set to 1 Hz and low-pass filter set to 40 Hz. Spectral data were smoothed using Hanning window techniques. Subject's eyes were closed producing a marked increase in power in the alpha (10 Hz) region, visible in both plots.

Research conducted to date suggests that no special skills or individual characteristics are required to develop the ability to self-regulate one's EEG. It is not unusual for new users to develop discernible control in the first training session. Proficiency at rudimentary one-dimensional control tasks typically requires 3-6 hours of training.

As with EMG, multiple EEG channels have been employed to increase the DOFs being controlled. For example, lateral control of a computer cursor has been based on the difference in 10 Hz power between the two brain hemispheres, while vertical control was based on the sum of this power in the two hemispheres.

Another approach to EEG-based control is based on the detection of spontaneous EEG patterns associated with the preparation for specific body movements, eye fixations or utterances. These approaches require no user training, but often require training of the pattern recognition algorithms. The number of training repetitions has ranged from approximately 100 to 1000 when neural networks (see Appendix C) have been employed as pattern recognizers. As one might imagine, the amount of time required for neural network training is highly dependent on the specificity of the data being sent to the network, with highly discriminatory data requiring far less training.

Both EMG- and EEG-based control systems require much more calibration and individual tuning than do conventional controllers. For example, in the EEG self-regulation approaches, signal threshold and duration requirements can and should be modified as users develop better control of their EEG. Modifying these parameters permits better discrimination of intentional control from background EEG activity. While research applications commonly use a trained operator to perform this calibration and tuning, ongoing work suggests that much of this can be automated. EEG and EMG control systems are typically calibrated by first measuring the dynamic range of the users signal and defining switching levels based on these measurements. More complicated systems based on pattern classification techniques rely on acquiring signal feature templates that may be different for each operator. Fortunately, it does not appear that there is a great deal of variation in daily calibration values of either EEG- or EMG-based systems, once users have completed the training process.

2.5.2 OVERVIEW OF APPROACHES

In this section we will review the methods used to acquire and process EMG and EEG signals. Of necessity, some information on the specific control applications will be included in the methodology descriptions. Nevertheless, details on the applications and on user performance with these systems will be reserved for section 2.5.3.

2.5.2.1 Signal Acquisition

Although implanted electrodes continue to be explored for some biopotential control applications, it is unlikely that they will be employed in near-term aerospace environments. EMG and EEG signal acquisition is most commonly accomplished using metal, coated plastic or gel electrodes located on the surface of the skin. Mild cleaning of the skin is often performed to reduce the impedance of the electrode-skin interface. EMG electrodes are usually applied dry and rely on high input impedance amplifiers and the development of a perspiration layer to reduce common mode interference. EEG electrodes are commonly applied with a conductive paste or cream and affixed with adhesive rings, tape or an elastic band. Gel electrodes do not require a conductive paste since the gel itself contains an electrolyte. Aerospace applications will benefit from convenient dry electrode systems, but these are not yet commercially available for EEG recording. Bandpass and notch filters are commonly employed to eliminate DC drift, AC line noise and to focus on the signal frequency range of interest. Commercially available biological signal amplifiers are well suited to the amplification and filtering of both EMG and EEG signals.

The next signal acquisition step in most biopotential controllers is analogue to digital signal conversion. Signal processing is most commonly performed in the digital domain although analogue processing is sufficient for simple amplitude based systems. Personal computer systems, in some cases with digital signal processing boards added, provide sufficient computational power to implement the signal processing approaches reviewed below. Thus the size, cost and weight of biopotential-based control systems are not serious constraints.

2.5.2.2 Signal Processing

Two general approaches characterise most current EMG- and EEG-based control systems:

- The use of EMG and EEG responses, not normally associated with motor control, to operate external devices. For example, raising the EEG activity in a specific frequency band might be used to turn a switch on or off.
- The use of spontaneous EMG and EEG patterns, normally associated with sensory or motor activity, to produce a similar response in an external device. For example, the remaining movement-related myoelectric activity in the arm of an amputee might be used to operate a prosthetic hand.

Each of these approaches is discussed separately for EMG- and EEG-based systems, below.

2.5.2.2.1 Self-Regulation of EMG Responses

To enable the EMG signal to be used as a control means, some feature of the signal must be extracted and an association must be made between values of this feature and the desired control response. The simplest EMG feature which can be extracted is signal amplitude. However, due to the random nature of the underlying myoelectric signal generation process the average value of the EMG signal is zero. Consequently, any attempt to filter the EMG signal to produce a smooth output for control purposes will result in a zero signal. To remedy this problem the signal must be processed to produce a signal which reflects the variance of the EMG signal. Although a square law device has been shown to be the optimum processor based on error probability [2.5-9], most controllers approximate this non-linearity using a full-wave rectifier. By amplifying, rectifying and filtering the EMG signal, a control signal can be obtained based on the effort of the voluntary muscle activity. Most current EMG-based control systems use an approach based on this simple amplitude feature.

Several types of control algorithms have been developed. These can be divided into three general categories: (a) Level

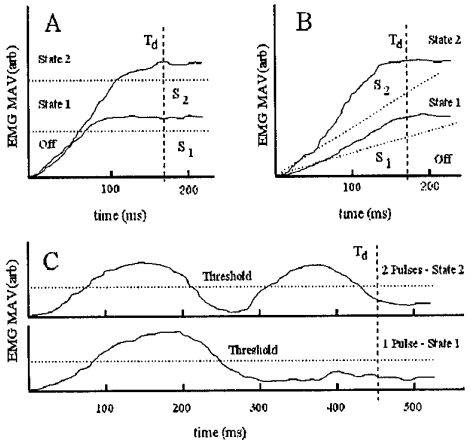


Figure 2.5-3 EMG Signal Coding: (A) Level, (B) Rate, (C) Pulse. MAV = Mean Absolute Value.

coding, (b) Rate coding and (c) Pulse coding.

Level Coding. To control a single degree-of-freedom device (i.e., prosthetic hand) the control signal can be derived from either the EMG signal level of a single muscle or from the two EMG signal levels from an agonist/antagonist muscle pair (i.e., flexor/extensor). In the first case, the dynamic range of the signal (the range between the noise and the maximum EMG signal produced) is divided into three regions by two switching thresholds giving a 3-way switch to control the state of the terminal device (off, hand open, hand close) (Figure 2.5-3a). In the latter case, each signal controls one state switch (i.e., flexor EMG for hand open, extensor EMG for hand close). To avoid the situation where both switches are in the on position (i.e., co-contraction of the two muscles), control is given to the larger of the two signals or to whichever signal first exceeded the switching threshold.

Rate Coding. Rate coding works on the principle of how fast the user contracts the control muscle (Figure 2.5-3b). A control signal is derived based on the initial slope of the processed EMG signal from a single electrode site. A slope threshold is set such that a slow contraction selects one function (i.e., hand open) and a fast contraction selects another (i.e., hand close). The operator performance and training requirements are similar to the single channel level-coded system.

Pulse Coding. It is also possible to derive a control signal based on pulses of EMG activity (Figure 2.5-3c). A simple coding scheme can be devised to define the control output signal. Function selection is then just a matter of producing the associated pulse code (i.e., one pulse - hand open, two pulses - hand closed).

Advantages and Limitations of each Approach. Clinical experience has shown that control systems based on either level coding or rate coding are easy to operate and that operator error is insignificant (Figure 2.5-4a) after a short period of training [2.5-10]. However, if the dynamic range is segmented into more than three regions (Figure 2.5-4b), in an effort to extract more control information, the operator error increases quickly [2.5-11]. Gains and switching level settings for each system depend on the individual's EMG levels and must be adjusted to achieve optimum control. Further adjustments may be required during the initial training period but little adjustment is required thereafter.

For both level-coded and rate-coded systems the operator does not notice the small time delay introduced by the control system. A system based on pulse coding, however,

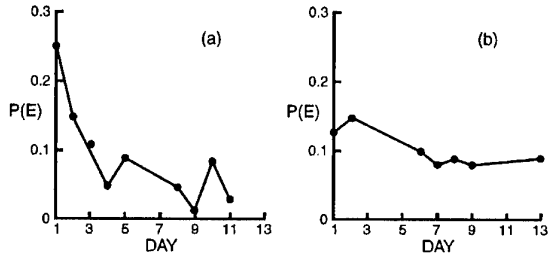


Figure 2.5-4 Training Effect on the Probability of Operator Error P(E) for (a) 3-State and (b) 5-State Amplitude-Coded EMG Control Systems. From [2.5-10]

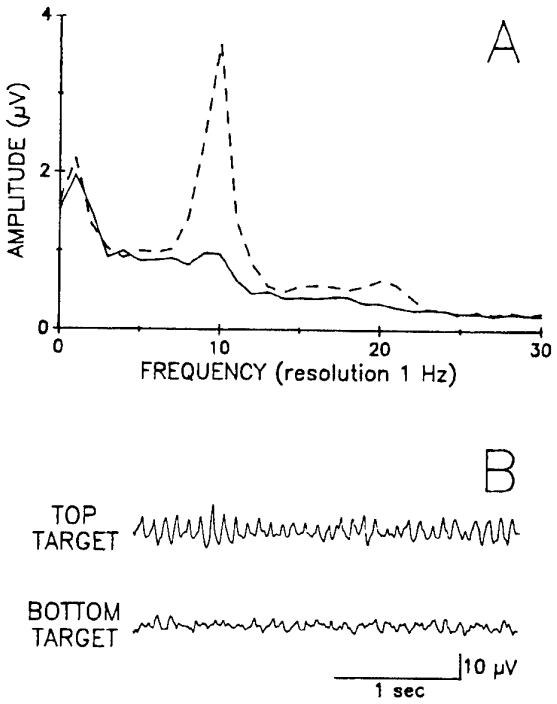


Figure 2.5-5 EEG Rhythm Level Coding. High power in the 10 Hz region moves a cursor toward the top target and low power moves the cursor toward the bottom target. (A) Frequency spectrum of EEG signal for top (dashed line) and bottom (solid line) targets, (B) sample EEG traces. From [2.5-13].

introduces a noticeable delay due to the operator's inability to produce rapid EMG pulses. Although this delay limits the application of this type of control in prosthetics, this form of system coding does have the potential to control a large number of devices.

The method used to control the speed and direction of the selected function is determined by whether there is a single EMG control channel or if two sites can be found to give two independent EMG control channels. A single channel system requires two separate functions to be assigned to a device to allow it to be controlled in both directions. Additional processing on the EMG control signal is necessary to determine which of the two functions is to be selected. In this case the device function is usually controlled in a simple on/off manner at a predetermined rate until the EMG control signal associated with that function is no longer detected. This means that the operator must relax (to turn the function off) before an alternative function can be selected. If two EMG channels are available, then each of the two device functions can be assigned to a separate channel. The device can then be driven at a speed proportional to the level of the EMG control signal and in the direction determined by the assigned channel. This proportional control is preferable to the state control of the single channel systems since corrections in position can be made more quickly.

2.5.2.2.2 Self-Regulation of EEG Responses

EEG Rhythm Level Coding. Level-coding techniques have been employed in several examples of EEG-based control. The EEG amplitude in a specific frequency band is

determined using fast Fourier analysis, bandpass filtering, or some other technique, and this amplitude is compared to set threshold criteria or used as the input variable in a linear equation. For example, small amplitudes might move a computer cursor downward, medium amplitudes produce no motion, and large amplitudes might move the cursor upward (Figure 2.5-5). Alternatively, simple linear equations have been employed to control cursor movement [2.5-12]. With an intercept of 4 microvolts and a slope of 5, a 3 microvolt signal would move the cursor down 5 steps, a 5 microvolt signal would move it up 5 steps, and a 6 microvolt signal would produce a 10 step upward output. Level coding has also been applied to signals from multiple EEG channels to increase the DOFs controlled. For example, the sum of the signals from the two hemispheres has been used to control vertical cursor motion, while their difference controlled horizontal movement [2.5-12].

Evoked Response Level Coding. Level coding of the amplitude of externally-evoked, as opposed to internally-generated, EEG signals has also been successfully employed. In this case, the brain response is produced by an external stimulus, such as a flickering light. With biofeedback and training, users can learn to modulate the amplitude of the brain's response to such stimuli. One implementation [2.5-14] produced a discrete control output when the amplitude of the evoked response remained above an experimenter-specified threshold for 75% of the samples in a one-half second interval. This combination of threshold and duration criteria required the user to produce sustained changes in the response; however, brief fluctuations did not interrupt system control. Further information and a figure illustrating this approach are included in Section 2.5.3.2.

Advantages and Limitations of Each Approach. The principal difference between the two EEG self-regulation methods discussed above is the presence or absence of an evoking stimulus. In both approaches the user controls the amplitude of a brain signal, but in one case the fundamental signal is evoked by external events. The use of an evoking stimulus complicates the interface design and requires that some of the user's sensory and perceptual resources be devoted to the processing of this input. In addition, the evoking stimulus may serve as a distraction or be poorly accepted by some users. On the other hand, the evoking stimulus produces a time-locked EEG response. This permits one to use synchronous signal processing techniques that improve noise tolerance and reduce or eliminate the confounding effects of other activities and rhythms. An open question concerning the self-regulation of internally-generated brain rhythms is the applicability of this approach with active, multitasked users; how difficult will it be to discriminate intentional and natural variation? At the present time, it is premature to discount either of these approaches based upon such considerations. Only further development and application will identify the real constraints associated with each method.

Both of the self-regulation approaches require significant calibration or adjustment for individual users early in the training process. Once users establish reliable control, the calibration values tend to remain quite stable from day to day. The use of linear equations in some of the EEG controllers does not mean that users can precisely control their EEG amplitudes. In fact, it appears that users can reliably produce signals in only 2-3 broad amplitude ranges. Nevertheless, the use of linear equations does permit a rudimentary form of proportional control.

2.5.2.2.3 Interpretation of Spontaneous EMG Responses

Recent approaches to EMG control are based upon the interpretation of spontaneous EMG signals associated with natural muscle contractions. The impetus for such a system is that it would require little or no user training. No longer is the operator required to produce somewhat unnatural, self-regulated contractions. The control system learns to recognise the spatial and temporal patterns within the EMG signals, from one or several muscles, during contractions which correspond naturally to the desired controlled function. For example, an above-elbow amputee may choose to train the control system to associate the patterns produced during stump rotation with selection of wrist control. In other words, the training function is shifted from the operator to the control system.

Pattern Recognition. All EMG-based control systems implemented using pattern recognition have been based on the assumption that at a given electrode location, the set of

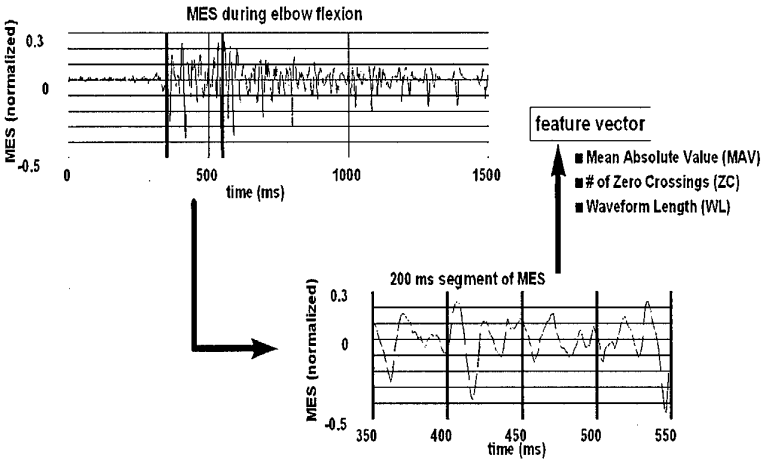


Figure 2.5-6 Representation of the EMG Pattern by a Time Feature Vector. MES = Myoelectric Signal.

parameters describing the EMG will be repeatable for a given state of muscle activation and furthermore that it will be different from one state of activation to another. To control ***m*** distinct remote functions requires ***m*** unique patterns of activity. Control schemes have been based almost entirely on the discriminant approach to pattern recognition, in which each pattern is described by a set of signal features. These features may be EMG from a number of electrodes, a set of statistics describing the signal sampled at one electrode site, or some other reproducible set of features. Once the patterns are described in this feature space, an unknown pattern can be compared with them to determine which of the ***m*** functions should be selected.

The classical approach to this problem is that of Wirta [2.5-15] in which the activity (simply muscle active = on / muscle inactive = off) at a number of muscle sites is monitored and function activation is controlled on the basis of a match between the observed activity and a predefined on/off pattern across all sites. Graupe [2.5-16] has suggested a somewhat different approach in which the signal observed at a single site is defined by a set of coefficients which are indicative of the activity of nearby muscles. These coefficients vary depending on the type of contraction being produced, due to the involvement of different muscles. These coefficient patterns are treated in the same way as the on/off patterns from different sites, for control of several functions.

A recent example uses a neural network to classify specific patterns in the EMG signal from natural voluntary contractions of the residual limb [2.5-17]. In this example the pattern classifier is trained to recognise the specific contractions based on a set of time domain features. The features are extracted from a single EMG signal during reproductions of several contraction types. The classifier then uses this information to develop a feature template or signature for each contraction type (Figure 2.5-6). During use, each contraction produced by the operator is compared to all templates to determine which is most similar. The control system then switches to the device which corresponds to this choice. As in the examples of Section 2.5.2.2.1, this again is a simple selection scheme in which one of four states of the prosthetic limb is selected to be controlled. This research has been extended to a more functional system which uses the same classifier structure to control a three DOF prosthetic limb [2.5-4]. In this system the feature information is extracted from two independent EMG signals (i.e., flexor/extensor) which allows control of the speed and direction of the selected DOF.

Advantages and Limitations of Each Approach. The key advantage of a control system based on pattern recognition is that the training burden is moved from the operator to the control system. This assumes, however, that the operator will produce patterns which are unambiguous to the classifier. It is often the case that only a small number of distinct patterns can be found. Each new pattern class entered into the training set reduces the available feature space and increases the chance of pattern class overlap.

Pattern recognition systems based on on/off muscle activity patterns require many channels of EMG amplifiers and signal conditioners and a large number of accessible muscle sites. This is reduced in the single- and dual-channel systems, however, more complex signal processing hardware and software is necessary to achieve comparable system performance. The recognition of on/off events can also be done very quickly. Systems based on the recognition of more complex time or frequency domain features require an analysis of a much longer sample of the EMG signal to reduce the feature estimation error . This can introduce a noticeable time delay in the selection process.

2.5.2.2.4 Interpretation of Spontaneous EEG Responses

Several approaches to EEG-based control are based upon the interpretation of spontaneous brain responses. Using this approach, little or no user training is required. Informal observations suggest, however, that overall system performance may improve with experience, i.e., users may develop the ability to enhance their spontaneous responses in order to improve their control.

Evoked Response Level Coding. Natural variation in the amplitude of brain responses evoked by external stimuli can be used for control. One example is the P300 component of the event-related potential (ERP). The P300 is a positive-going component of the ERP response to a sensory input, with an amplitude in the 5-10 microvolt range and a latency of about 300 milliseconds. This response is most prominent over the central and posterior (parietal) regions of the scalp. Many studies have demonstrated that the P300 is enhanced when subjects are presented with a stimulus that is of low probability or has special significance. If, for example, a user is asked to select a particular item that is presented in a series of items, the user will produce a larger P300 when the desired item is presented [2.5-18].

Unfortunately it is not possible to reliably discriminate among P300 responses to single presentations of a series of items. Multiple presentations and response averaging are required. Farwell and Donchin [2.5-19] investigated the rate of stimulus presentation, the number of presentations, and the type of signal processing algorithm while using the P300 response as a means for subjects to select one element from a 36 element matrix of letters and words. In this case the stimuli were repeated intensifications of the rows and columns of the matrix. Farwell and Donchin found that they could discriminate among the P300 responses using interstimulus intervals as short as 125 milliseconds, which caused the responses to overlap. Despite this overlap, a minimum of 26 seconds was required to generate multiple, separable responses to each of the 36 matrix elements. They compared stepwise discriminant analysis, response peak picking, response area computation, and covariance computation approaches for P300 amplitude identification. The best method varied from subject to subject and was also dependent on the rate of stimulus presentation.

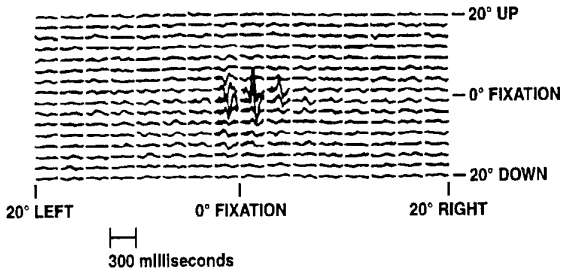


Figure 2.5-7 VEP to Central and Peripheral Modulated Visual Stimuli. From [2.5-20].

Another spontaneous response that has been evaluated for EEG-based communication is the visual evoked potential or VEP. This microvolt-level signal is produced by visual stimuli such as flashes and colour reversals. The major components of the transient VEP occur within 80 milliseconds of the stimulus, and are most commonly measured over the posterior (occipital) region of the scalp. As with the P300 response, multiple presentations and response averaging are required to estimate the amplitude of the VEP.

Sutter [2.5-20, 2.5-21] used the VEP as a means for subjects to select elements from an 8 by 8 matrix of letters and words presented on a computer monitor. The matrix elements were modulated in intensity or colour and the VEP to each modulated element was individually computed. His approach was based on the fact that a modulated stimulus in the centre of visual field evokes a much larger VEP than one in the visual periphery (Figure 2.5-7). The system selected the character with the largest response as the desired one, i.e., the one the user was visually fixating.

A powerful methodological aspect of Sutter's approach was the use of m-sequences (white pseudo-random binary sequences) to control the elements of the flickering matrix and to extract the average response to each matrix element from the combined signal (which included hundreds of overlapping VEP responses). The use of m-sequences allowed Sutter to generate multiple, separable responses to each of the 64 elements in approximately 1.5 seconds. Each of these responses was correlated with a reference VEP template collected in a 10-20 minute preliminary session. These correlation coefficients were then compared to each other and to a threshold value. If coefficient n remained above threshold and was larger than all others for a specified amount of time, then matrix element n was selected.

Pattern Recognition. Rather than focusing on the amplitude of a single EEG response, control can be based on more complex spectral, temporal or spatial patterns in the EEG. For example, back-propagation neural networks have been trained to recognise snapshots of the EEG amplitude, at multiple electrode locations, sampled just prior to uttering a vowel or moving a joystick [2.5-22]. With such approaches, as many as 1000 repetitions of each utterance or joystick movement may be required to train the neural network.

Alternatively, one may focus on more specific patterns of EEG activity associated with cortical preparation for body movements. One such pattern is the reduction in mu rhythm (8-12 Hz) power in the sensorimotor area of the cerebral hemisphere contralateral to the movement [2.5-23, 2.5-24]. Pfurtscheller and his colleagues have attempted to use this and other such patterns to classify finger, toe or tongue movements before they actually occur [2.5-25, 2.5-26]. In particular, they have focused on power decreases (event related desynchronization, or ERD) in the 8-12 Hz band and brief power increases (event related synchronisation, or ERS) in 30-40 Hz band (Figure 2.5-8).

These signals were recorded with an array of 8-14 electrodes spread over the central and parietal regions of the scalp and processed using DSP techniques. Classification was accomplished with Kohonen Learning Vector Quantization (a type of neural network) [2.5-27] which iteratively defines a set of reference vectors for each classification category, in this case finger, toe or tongue movement. Following training, these reference vectors were used by the network to classify new EEG input vectors. Pfurtscheller and his colleagues typically used only 100-200 trials to train the network, far less than the number employed by Hiraiwa. The greater temporal and spatial specificity of the EEG patterns being classified by Pfurtscheller may be a major contributor to reduced network training time.

While Hiraiwa and Pfurtscheller provided their neural networks with samples of EEG collected at successive time points, their static neural networks did not actually assimilate the temporal dimension of the patterns being classified. With static networks, the successive samples are provided as simultaneous inputs to separate nodes of the input layer. Barreto, Taberner and Vicente [2.5-28], have recently begun to evaluate the potential of dynamic neural networks for the classification of EEG patterns which represent preparation for body movements. With dynamic neural networks, the input consists of a temporal sequence of values provided to a single input node. Such networks store past samples of the inputs in memory structures that perform a time-to-space mapping for the classifier.

Advantages and Limitations of Each Approach. The evoked response approaches require no training of the user, or of the signal processing algorithms. Since the responses are essentially time-locked to an external stimulus, selecting temporal windows for signal processing is readily accomplished. However, the requirement to average multiple responses, in order to obtain reliable amplitude estimates, can be a significant constraint. Nearly sequential presentation of the evoking stimuli limited Farwell and Donchin [2.5-19] to very low character selection rates. Sutter's [2.5-21] use of m-sequences permitted highly-overlapping stimulus presentation and significantly improved the output bandwidth of his system.

The pattern recognition techniques, which all employ neural

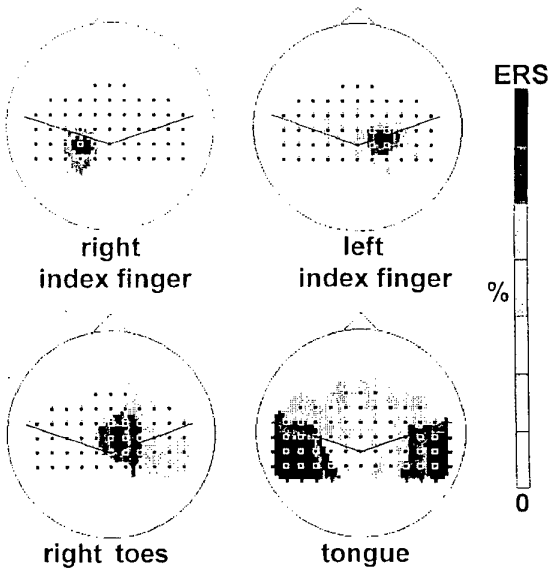


Figure 2.5-8 Brain Maps Showing Specific Patterns of Event Related Synchronisation (ERS) Associated with the Preparation for Specific Body Movements. From [2.5-26].

networks, require individual training of the recognizer. As noted above, the neural network must be trained with 100-1000 repetitions of the EEG patterns to be classified. One potential approach to this training issue is to conduct it implicitly rather than explicitly. For example, the user might continue to physically operate the controls, while the EEG-based recognizer observes the brain patterns associated with these activities. Once the recognizer can satisfactorily predict certain control actions, it could then be permitted to take over those functions. This implicit training might be conducted in simulated or synthetic environments, for example.

Practical application of the pattern recognition techniques must address the issue of selecting the temporal window that includes the EEG patterns to be recognised. This problem is analogous to the challenge faced by speech recognizers operating in a high noise environment. The experiments conducted to date create an artificial solution to this problem. The user is given explicit cues to execute the movements to be predicted from the EEG, and the pattern recognizer is synchronised to these cues. In the real world, such cues typically will not exist. Barreto et al. [2.5-28] argue that dynamic neural nets will reduce this problem, but this advantage has not been demonstrated in a real world environment.

Finally, individual differences represent an additional constraint on both approaches. Evoked response amplitudes, dynamic EEG patterns, artefact characteristics and optimal electrode locations all vary from person to person. The pattern recognition techniques tend to address these issues during recognizer training, while the evoked response approaches often require initial tuning of electrode locations, signal processing algorithms and response templates for each individual. Fortunately, many of these sources of variation are fairly stable from day to day, once the signal processing parameters have been optimised for each user.

One can also compare the approaches based on EEG self-regulation with those that employ spontaneous EEG responses (section 2.5.2.2.2 above). The former are clearly less natural and intuitive than the latter, since the user must produce artificial changes in their EEG. This does not mean that such changes are inappropriate, interfere with other cognitive activities, or are difficult to produce. Rather, users must learn to produce these changes, and this requires an investment in training. Once these EEG patterns are under voluntary control, there is a great deal of flexibility in how these patterns can be applied. Essentially, their application is constrained only by the bandwidth, resolution and accuracy of EEG self-regulation.

2.5.3 APPLICATIONS TO DATE

2.5.3.1 EMG

2.5.3.1.1 As a Stand-Alone Controller

EMG-based control is not new; the first system, which used a single channel rate-coding control strategy, was designed by Reiter in 1948 [2.5-29] and was used in a prosthetic arm to assist amputee factory workers in performing assembly line tasks. However, commercial systems did not become widely available until the 1970s due to size and power consumption limitations. EMG-based systems are now used extensively as controllers of prosthetic devices for individuals with amputations or congenitally deficient upper limbs, and as augmentative controls for the severely disabled. Many systems are now available commercially to control a single prosthetic device (e.g., hand, elbow, wrist). Most use either amplitude coding [2.5-30] or rate coding [2.5-31] to control this single DOF. Richard et al. [2.5-32] have implemented an amplitude-coding scheme in which the operators are given an indication of which state they are in by way of a state dependent electric stimulus. This state feedback has the potential of increasing the number of DOFs controlled per EMG channel, however, the electrical stimulus and the increased vigilance required to use the system are not tolerated by all users. Feedback in the form of a force dependent electrical or mechanical stimulus has also been used to give the operator a sense of touch [2.5-33]. This force feedback improved the operator's grasping ability by replacing, in part, the lost sensory feedback of the natural hand, but it has not been widely accepted. The problem of controlling more than two states (one DOF) per EMG channel has been solved using approaches based on the recognition of EMG patterns. Many researchers [2.5-16, 2.5-34] have developed multifunction control systems based on the recognition of features extracted from the EMG patterns involved in normal muscle contraction. As expected, these systems have been shown to provide excellent selection

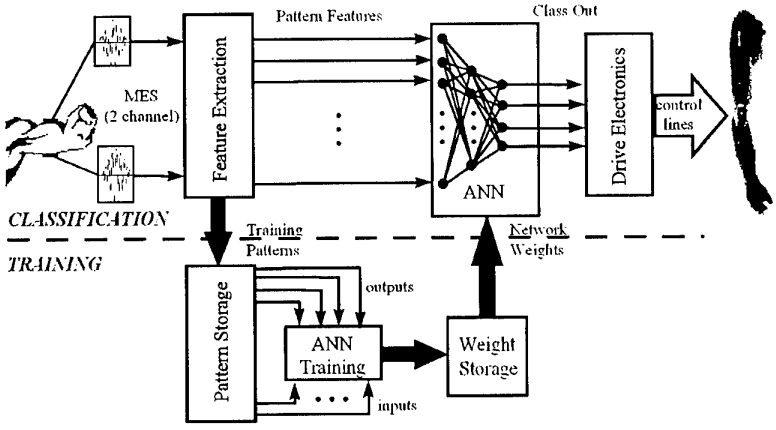


Figure 2.5-9 A Neural Network-Based Myoelectric Control System. MES = Myoelectric Signal.

accuracy with minimal user training. The most recent system, shown schematically in Figure 2.5-9, provides the amputee with good control of up to four DOF [2.5-4]. However, these systems are still experimental and none have achieved clinical acceptance.

Basic on/off EMG amplitude-based systems have been used to implement control schemes for electric wheel-chairs, orthotic aides, environmental systems and communication devices. EMG-based control systems have also been of interest to the able-bodied community. Early work sponsored by the US Air Force [2.5-35] showed that on/off EMG patterns from six sites on the upper arm could discriminate the purposeful muscular actions of pilots in simulated high-g environments. An EMG-based control system was designed which controlled the movement of a splint to provide a powered assist to the pilot's arm. The pilot was able to achieve 90% accuracy in a tracking exercise using this system.

The National Aeronautics and Space Agency (NASA) has sponsored several studies on the possibility of using EMG control for robotic teleoperation applications. Clark and Phillips [2.5-36] found that EMG time histories from normal hand and arm motion were not appropriate for controlling the complex movement kinematics of a robot arm. However, more recent work by Farry et al. [2.5-37] has found that a time-frequency analysis of the EMG patterns from forearm musculature could discriminate several different hand grasp types and thumb motions with a high degree of accuracy. Fernandez et al. [2.5-38] has continued this work using a classification scheme based on genetic programming (a neural network training technique) to achieved 100% classification of thumb motions from the same EMG data. These results suggest that it may be feasible to use EMG from an operator's own hand and arm to replace or augment joysticks and exoskeletal instrumentation as the master to intuitively control a remote anthropometric robot arm.

2.5.3.1.2 In Combination with Other Controls

It is common practice in clinical prosthetics to use EMG-based control in conjunction with mechanical switches to achieve a prosthetic limb in which the amputee can simultaneously control two DOF [2.5-39]. Typically, this system consists of an electric hand controlled by the EMG and a switch-controlled elbow.

Recent work by Junker, Berg, Schneider and McMillan [2.5-40] has shown that subjects can use a combination of EMG and EEG (referred to in their work as the brain-body signal) extracted from electrodes on the forehead to control the movement of a cursor to track computer-generated targets. Although the best tracking performance was achieved by subjects using mainly EMG control, there was evidence that other subjects scored better than chance using primarily an EEG-based response. This group has also found that, for discrete on/off responses, a brain-body actuated control scheme can achieve high classification accuracy with little user training and with reaction times comparable to manual switches [2.5-41]. Vodovnik [2.5-42] has shown that reaction time can be enhanced using an EMG trigger. That study showed a substantial reduction in reaction time when an electronic braking system, triggered by EMG from the frontalis muscle, was used to augment a normal foot-activated automobile brake.

The EMG signal has the potential to augment more traditional control methodologies. For example, recent work [2.5-43] has shown that phonetically-relevant orofacial motions can be estimated from the underlying EMG activity. It is reasonable to assume that information from the EMG of facial muscles could improve the performance of current speech recognition systems. There is also a possibility that information from neck and shoulder muscle EMG could aid in determining head position and orientation. Kang et al. [2.5-44] have reported a 86.7% success rate in classifying ten head and shoulder movements using EMG pattern information from the Trapezius and Sternocleidomastoid muscles.

2.5.3.1.3 For Operator State Monitoring

In addition to their use as a control signal, EMG signals have been used to indicate muscle activation state. Most biofeedback systems are based on monitoring the EMG signal to provide a visual or auditory representation of the level of muscle relaxation. The relationship between EMG and muscle force has been used extensively in ergonomic studies of many occupations, including pilots, to define workloads [2.5-45]. The shift in median frequency of the EMG signal has become accepted as a reliable measure of localised muscle fatigue [2.5-46].

2.5.3.2 EEG

To date, EEG-based control has not been integrated with other modalities, and all evaluations have been conducted in the laboratory. Nevertheless, a variety of interesting applications have been investigated.

2.5.3.2.1 As a Stand-Alone Controller

Computer Control. Several basic computer operations have been demonstrated with EEG-based control systems. Wolpaw and his colleagues [2.5-12, 2.5-13] have developed a system for computer cursor control that is based on voluntary modulation of the 8-12 Hz mu rhythm. Although it is in the same frequency range as the alpha rhythm, mu is recorded over the primary sensorimotor area of the brain and responds in known ways during movement preparation. In the single-axis task, the user moved the cursor to contact targets that appeared randomly at the top or bottom of the monitor. After approximately 18 hours of training, users required 2-6 seconds to move the cursor to a target. The target was correctly selected on 80-95 percent of the trials. The dual-axis task used mu rhythm signals from both cortical hemispheres in a more complex control algorithm. After approximately 12 additional hours of training, the users required 2-4 seconds to move the cursor to targets that appeared in one of the four corners of the screen. The target was correctly selected on 40-70 percent of the trials.

The work of Farwell and Donchin [2.5-19] and of Sutter [2.5-20, 2.5-21] represents a form of computer keyboard operation. In both cases a virtual keyboard was presented on the computer monitor and the user selected letters, words and numbers via EEG signals. Although Farwell and Donchin's use of the P300 response was interesting, the best performance that they were able to achieve was 2.3 characters per minute, which is not practical even for persons with severe disabilities. As detailed in Section 2.5.2.2.4, Sutter modulated the keyboard elements and determined which one produced the largest VEP. By using virtual keyboard overlays, communication rates of 10-12 words per minute were achieved. For example, the first keyboard overlay

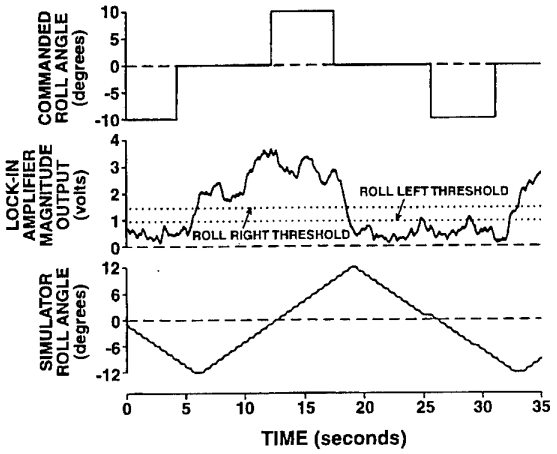


Figure 2.5-10 Control of Simulator Roll-Axis Motion Using the Visual Evoked Response. The lock-in amplifier provides a continuous measure of the magnitude of this response. Responses above one threshold produce motion to the right (positive), while responses below a lower threshold produce motion to the left (negative). From [2.5-48].

contained the alphabet and many frequently used words. If the desired word was not on that screen, the user selected the beginning letter, which brought up an overlay of words beginning with that letter. Clearly, Sutter's approach provides practical communication rates, but potential users must deal with a somewhat unpleasant flickering keyboard display. Since the VEP and P300 are natural responses to external stimuli, little or no user training is required to employ them as controls.

Finally, the use of an EEG signal as a mouse click has been demonstrated. In a manner similar to Wolpaw et al., LaCourse and Wilson [2.5-47] used the mu rhythm for single-axis cursor control. They employed the alpha rhythm as a means to switch between horizontal and vertical modes of control. In most individuals, a marked increase in alpha rhythm amplitude is observed when the eyes are closed (Figure 2.5-2). LaCourse and Wilson used this response as a mouse click to switch modes of the mu rhythm controller and to turn a device on or off when the cursor was positioned over the appropriate switch.

Control Selection and Activation. The work of Farwell and Donchin [2.5-19] and of Sutter [2.5-20, 2.5-21] can also be characterised as the selection and activation of discrete switches. Applied in this manner, performance equal to or better than virtual keyboard operation would be expected, depending on the number of switches to be discriminated.

Calhoun and McMillan [2.5-14] investigated discrete switch selection using self-regulation of the visual evoked response to a modulated light incorporated in the task display. In their experimentation, the switches represent modes of a fighter aircraft weapon system, and the operator must select the weapon mode based upon the range of the target. Cycling through the modes was initiated by holding the evoked response above an amplitude threshold. Once the desired mode was selected, the operator had to suppress their response to stop the cycling. Simulated weapon firing was then initiated with a trigger pull. After approximately 10 hours of training, operator accuracies ranged from 60-90% correct mode selection and average mode selection times ranged from 4-6 seconds.

Pfurtscheller and his colleagues [2.5-25, 2.5-26] are developing the technology to "automate" simple control actions normally performed with the hands or feet. Neural networks are employed to recognise mu rhythm and gamma band (30-40 Hz) EEG patterns that precede specific body movements, such as finger, toe or tongue activity. No user training is required, but the movements must be repeated 100-200 times for neural network training. After training, movement prediction is possible with only one second of EEG data. Pfurtscheller's off-line system achieved 89% accuracy in predicting button pushes with the left or right hand. With toe and tongue movement added, accuracy dropped to 70%. In addition, the neural networks can be trained with imagined rather than actual movements, but movement prediction is slightly degraded in this case.

Virtual joystick operation has been demonstrated by Hiraiwa, Shimohara and Tokunaga based upon neural network recognition of the EEG patterns that precede joystick movement [2.5-22]. Following network training, the authors were able to predict the direction of joystick movements with 96% accuracy. This evaluation was conducted with one subject and off-line analysis of the data. By way of comparison, the authors also attempted real-time prediction of the utterance of one of two Japanese vowels and reported 100% success after 1000 network training trials.

Physical Device Control. Using self-regulation of the visual evoked response, McMillan and Calhoun have investigated EEG-based control of a number of devices, including the roll-axis motion of a simple flight simulator [2.5-48]. A task display in the simulator provided a random series of commands requiring the operator to roll right or left to specific target angles. The operator accomplished this control by raising the evoked response above a high threshold to roll right and suppressing the response below a low threshold to roll left. A typical simulator control trial is shown in Figure 2.5-10. Typically, subjects were able to acquire 70-85% of the roll-angle targets after 5-6 hours of training (Figure 2.5-11).

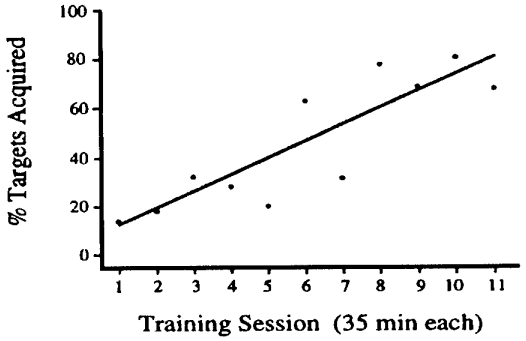


Figure 2.5-11 Learning Curve for One Subject Performing the Roll-Axis Motion Control Task using the Visual Evoked Response. Subject had no prior biofeedback or simulator training. Data points are means of 16 trials in each session. Solid line is a linear regression on these means. From [2.5-48].

In addition to the simulator control application, the same group employed evoked response control to operate a neuromuscular stimulator designed to exercise paralysed limbs [2.5-49]. Raising the evoked response above a high threshold turned the stimulator on and suppressing the response below a low threshold turned it off. This, in turn, caused the user's knee to extend or flex in response to changes in stimulator current. A series of specific knee extension angle commands was presented in each trial to test user performance. A group of three subjects was able to acquire 96% of the knee angle targets presented in a brief pilot study.

2.5.3.2.2 For Operator State Monitoring

A closely allied application of EEG measurement is operator state monitoring. In this case the EEG provides the system with on-line data, not otherwise available, about the user's level of mental workload and alertness [2.5-50, 2.5-51, 2.5-52, 2.5-53]. Research in this area has emphasised three domains of strategic importance to aerospace operations: operator workload monitoring, error prediction and monitoring, and fatigue monitoring. Gevins and his colleagues have applied neural network pattern recognition techniques to EEG data and found that it is possible to automatically determine whether or not momentary inattentiveness [2.5-54] or incipient fatigue resulting from sustained operations [2.5-55] is likely to lead to errors in behavioural performance. Wilson and Fisher [2.5-56] were highly successful in classifying which of 14 laboratory tasks subjects were performing. Principle components and stepwise discriminant analyses resulted in 86% classification accuracy using topographic EEG data. These and other state-monitoring advances hold promise for near-term transition into a variety of military and civilian sectors.

2.5.4 REQUIRED ENHANCEMENTS AND PROGNOSIS

This review demonstrates that a wide variety of tasks can employ biopotential signals for control. With creative interface design, almost any discrete response task can be performed with these modalities. In certain cases, response time advantages have been demonstrated using EMG signals to replace physical movement of a conventional control. The ability to perform continuous proportional control is clearly much more limited. Difficulty in producing and maintaining graded EMG and EEG signal outputs is the reason for this limitation. To achieve this type of control, investigators most commonly employ time-proportional techniques in which the position or velocity of the controlled element is proportional to the amount of time that a biopotential signal remains above an established threshold.

A clear limitation of the current state of the art, with the exception of prosthetic device control, is that little work has been done outside the laboratory. There is a profound need to identify applications which require hands-free operation and to develop biopotential controllers for field evaluation. Only then can developers determine how to achieve effective performance in real-world, multitask, multicontrol environments.

Improvements in signal acquisition hardware are required to support such applications. Dry electrode systems that do not require skin preparation or electrode creams are essential. These electrodes must work on hairy skin or scalp areas and be tolerant of slippage or movement. Self-administered user calibration approaches, as well as signal monitoring schemes that continually adjust the interface based on signal quality and background noise are required.

Signal processing enhancements are also needed in many areas. While discrete EMG and EEG responses can be identified rapidly, recognition of complex patterns is more time consuming, and constrains the speed of human-system interaction. Nevertheless, the pattern recognition approaches offer the greatest potential for discriminating signal from artefact and for controlling multiple DOF systems. In the area of EEG-based control, for example, it is apparent that complex pattern recognition will be the method of choice if we are to develop true "thought-based" interfaces.

Finally, as discussed by McMillan, Eggleston and Anderson [2.5-57], biopotential control may be most applicable when used in a manner that bridges the space between operator state monitoring and explicit control. Referred to as an intelligent controller paradigm, this approach employs an intelligent interpreter that monitors a range of human outputs, including EMG and EEG signals, to infer intent, a desire for information, etc. The interpreter then issues commands to the system, consistent with user intentions. Recognition of the EEG patterns which precede specific physical movements is a simple example of this notion. Detection of a specific EEG response permits the interpreter to infer that the user desires to push a specific button. The intelligent interpreter approach, rather than simple substitution of EMG and EEG signals for conventional inputs, represents the path to achieve optimal utilisation of biopotential-based control.

2.5.5 REFERENCES

2.5-1 Parker, P.A., and Scott, R.N., "Myoelectric Control of Prosthesis", CRC Critical Reviews in Biomedical Engineering, 13, Issue 4, 1986, pp 283-310.

2.5-2 Jacobsen, S., Knutti, D., Johnson, R., and Sears, H., "Development of the Utah Arm", IEEE Transactions on Biomedical Engineering, BME-29, 4, 1982, pp 249-269.

2.5-3 Hudgins, B., Parker, P., and Scott, R.N., "Control of Artificial Limbs Using Myoelectric Pattern Recognition", Medical and Life Sciences Engineering, 13, 1994, pp 21-38.

2.5-4 Parker, P., Scott, R.N., Hudgins, B., Hruczkowski, T., Hayden, J., and Englehart, K., "Coordinated / Simultaneous Control of a Multifunction Myoelectric Prosthesis", Progress Report #2 on NSERC Project No. CRD151174, April 1995.

2.5-5 Saridis, G., and Stephanou, H., "A Hierarchical Approach to the Control of a Prosthetic Arm", IEEE Transactions on Systems Man and Cybernetics, SMC-7, 6, 1977, pp 407-420.

2.5-6 Dr. Lisa Dolev, Personal Communication, 1996.

2.5-7 United Kingdom Patent Number GB 2 274 396 B, December 1996.

2.5-8 Taheri, B., Smith, R.L., and Knight, R.T., "A Dry Electrode for EEG recording", Electroencephalography and Clinical Neurophysiology, 90, 1994, pp 376-383.

2.5-9 Hogan, N., and Mann, R.W., "Myoeletric Signal Processing: Optimal Estimation Applied to Electromyography - Part 1: Derivation of the Optimal Myoprocessor", IEEE Transactions on Biomedical Engineering, BME-27, 7, 1980, pp 282-295.

2.5-10 Scott, R.N., Paciga, J., and Parker, P., "Operator Error in Multistate Myoelectric Control Systems", Medical and Biological Engineering, 16, 1978, pp 296-301.

2.5-11 Paciga, J., Richard, P., and Scott, R.N., "Error Rate in Five-State Myoelectric Control Systems", Medical and Biological Engineering, 18, 1980, pp 287-290.

2.5-12 Wolpaw, J.R., and McFarland, D J., "Multichannel EEG-Based Brain-Computer Communication", Electroencephalography and Clinical Neurophysiology, 90, 1994, pp 444-449.

2.5-13 Wolpaw, J.R., McFarland, D.J., Neat, G.W., and Forneris, C.A., "An EEG-Based Brain-Computer Interface for Cursor Control", Electroencephalography and Clinical Neurophysiology, 78, 1991, pp 252-259.

2.5-14 Calhoun, G.L., and McMillan, G.R., "EEG-Based Control for Human-Computer Interaction", in "Proceedings of the Third Annual Symposium on Human Interaction with Complex Systems", IEEE Computer Society Press, 1996, pp 4-9.

2.5-15 Wirta, R.W., Taylor, D.R., and Findley, F.R., "Pattern Recognition Arm Prosthesis: A Historical Perspective - Final Report", Bulletin of Prosthetics Research, 10-30, 1978, pp 9-35.

2.5-16 Graupe, D., Salahi, J., and Kohn, K.H., "Multifunction Prosthesis and Orthosis Control via Microcomputer Identification of Temporal Pattern Differences in Single-Site Myoelectric Signals", IEEE Transactions on Biomedical Engineering, BME-29, 4, 1982, pp 17-22.

2.5-17 Hudgins, B., Parker, P., and Scott, R.N., "A New Approach to Multifunction Myoelectric Control", IEEE Transactions on Biomedical Engineering, BME-40, 4, 1993, pp 82-94.

2.5-18 Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.M., and Gratton, G., "Cognitive Psychophysiology and Human Information Processing., in Coles, M.G.H., Donchin, E., and Porges, S.W., (Eds) "Psychophysiology: Systems, Processes, and Applications", New York, Guilford Press, 1986.

2.5-19 Farwell, L.A., and Donchin, E., "Talking Off the Top of Your Head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials", Electroencephalography and Clinical Neurophysiology, 70, 1988, pp 510-523.

2.5-20 Sutter, E.E., "The Visual Evoked Response as a Communication Channel", in "Proceedings: IEEE Symposium on Biosensors", IEEE, 1984, pp 95-100.

2.5-21 Sutter, E.E., "The Brain Response Interface: Communication through Visually-Induced Electrical Brain Responses", Journal of Microcomputer Applications, 15, 1992, pp 31-45.

2.5-22 Hiraiwa, A., Shimohara, K., and Tokunaga, Y., "EEG Topography Recognition by Neural Networks", IEEE Engineering in Medicine and Biology, 19, 1990, pp 39-42.

2.5-23 Jasper, H.H., and Penfield, W., "Electrocorticograms in Man: Effect of the Voluntary Movement upon the Electrical Activity of the Precentral Gyrus", Arch. Psychiat. Z. Neurol., 183, 1949, pp 163-174.

2.5-24 Chatrian, G.E., Petersen, M.C., and Lazarete, J.A., "The Blocking of the Rolandic Wicket Rhythm and Some Central Changes Related to Movement", Electroencephalography and Clinical Neurophysiology, 11, 1959, pp 497-510.

2.5-25 Pfurtscheller, G., Flotzinger, D., Mohl, W., and Peltoranta, M., "Prediction of the Side of Hand Movements from Single-Trial Multi-Channel EEG Data using Neural Networks", Electroencephalography and Clinical Neurophysiology, 82, 1992, pp 313-315.

2.5-26 Pfurtscheller, G., Flotzinger, D., and Neuper, C., "Differentiation between Finger, Toe and Tongue Movement in Man Based on 40 Hz EEG", Electroencephalography and Clinical Neurophysiology, 90, 1994, pp 456-460.

2.5-27 Kohonen, T., "The Self-Organizing Map", Proceedings of the IEEE, 78, 1990, pp 1464-1480.

2.5-28 Barreto, A.B., Taberner, A.M., and Vicente, L.M., "Classification of Spatio-Temporal EEG Readiness Potentials towards the Development of a Brain-Computer Interface", in "Proceedings of the 1996 IEEE SouthEastcon Conference", IEEE, 1996, pp 100-103.

2.5-29 Reiter, R., "Eine Neue Elektrounsthand", Grenzgebiete der Medizin, 4, 1948, pp 133-135.

2.5-30 Dorcas, D., and Scott, R.N., "A Three State Myoelectric Control", Medical Biological Engineering, 4, 1966, pp 367-372.

2.5-31 Childress, D., "A Myoelectric Three State Controller Using Rate Sensitivity", in "Proceedings of ACEMB", 1969, pp S4-S5.

2.5-32 Richard, P., Gander, R., Parker, P., and Scott, R.N., "Multistate Myoelectric Control: The Feasibility of 5-State Control", Journal of Rehabilitation Research and Development, 20, 1983, pp 84-86.

2.5-33 Phillips, C.A., "Sensory Feedback Control of Upper- and Lower-Extremity Motor Prosthesis", CRC Critical Reviews in Biomedical Engineering, 16, 1988, pp 105-140.

2.5-34 Saridis, G.N., and Gootee, T.P., "EMG Pattern Classification for a Prosthetic Arm", IEEE Transactions on Biomedical Engineering, BME-29, 6, 1982, pp 403-41.

2.5-35 Sullivan, G., Martell, C., Weltman, G., and Pierce, D., "Myoelectric Servo Control", Report to US Air Force Aeronautical Systems Division under Contract #AF33(657)-7771, May, 1963.

2.5-36 Clark, J.E., and Phillips, S.J., "The Efficacy of Using Human Myoelectric Signals to Control the Limbs of Robots in Space", NASA-CR-182901, 1988.

2.5-37 Farry, K.A., Walker, I.D., and Baraniuk, R.G., "Myoelectric Teleoperation of a Complex Robotic Hand", IEEE Transactions on Robotics and Automation, 12, 5, 1996, pp 775-778.

2.5-38 Fernandez, J., Farry, K.A., and Cheatham, J.B., "Waveform Recognition Using Genetic Programming: The Myoelectric Signal Recognition Problem", Genetic Programming 1996 Conference, 1996.

2.5-39 Williams, T.W. "Practical Methods for Controlling Powered Upper-Extremity Prostheses", Assistive Technology, 2, 1, 1990, pp 3-18.

2.5-40 Junker, A., Berg, C., Schneider, P., and McMillan, G.R., "Evaluation of the CyberLink Interface as an Alternative Human Operator Controller", US Air Force Technical Report AL/CF-TR-1995-0011, 1995.

2.5-41 Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Roe, M.M., Lu, L.G., Haas, M.W., Dennis, L.B., Pick, H.L., Junker, A., and Berg, C.B., "Brain-Body Actuated Control: Assessment of an Alternative Control Technology for Virtual Environments", in "Proceedings of the 1996 Image Conference", 1996, pp 225-232.

2.5-42 Vodovnik, L. "An Electromagnetic Brake Activated by Eyebrow Muscles", Electronics Engineering, 1967, pp 694-695.

2.5-43 Vatikiotis-Bateson, E., Munhall, K.G., Kasahara, Y., Garcia F., andYehia H., "Characterizing Audiovisual Information During Speech", in "Conference on Spoken Language Processing - CDROM", 1996, Paper No. 1010.

2.5-44 Kang, W., Cheng, C., Lai, J., Shiu, J., and Kuo, T., "A Comparative Analysis of Various EMG Pattern Recognition Methods", Medical Engineering Physics, 8, 5, 1996, pp 390-395.

2.5-45 Harms-Ringdahl, K., Ekholm, J., Schuldt, K., Linder, J., and Ericson, M., "Assessment of Jet Pilots' Upper Trapezius Load Calibrated to Maximal Voluntary Contraction and a Standardized Load", Journal of Electromyography and Kinesiology, 6, 1, 1996, pp 67-72.

2.5-46 DeLuca, C.J., "Myoelectric Manifestations of Localized Muscular Fatigue in Humans", CRC Critical Reviews in Biomedical Engineering, 11, 1984, pp 251-279.

2.5-47 LaCourse, J.R., and Wilson, E.W., "BRAINIAC: A Brain-Computer Interface", Instrumentation and Measurement Society Newsletter, 1996, pp 9-14.

2.5-48 McMillan, G.R., Calhoun, G.L., Middendorf, M.S., Schnurer, J.H., Ingle, D.F., and Nasman, V.T., "Direct Brain Interface Utilizing Self-Regulation of the Steady-State Visual Evoked Response", in "Proceedings of the RESNA 18th Annual Conference", 1995, pp 693-695.

2.5-49 Calhoun, G.L., McMillan, G.R., Morton, P.E., Middendorf, M.S., Schnurer, J.H., Ingle, D.F., Glaser, R.M., and Figoni, S.F., "Functional Electrical Stimulator Control with a Direct Brain Interface", in "Proceedings of the RESNA 18th Annual Conference", 1995, pp 696-698.

2.5-50 Gevins, A., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., Zhang, J., and Libove, J., "Towards Measurement of Brain Function in Operational Environments", Biological Psychology, 40, 1995, pp 169-186.

2.5-51 Humphrey, D., and Kramer, A.F., "Toward a Psychophysiological Assessment of Dynamic Changes in Mental Workload", Human Factors, 36, 1994, pp 3-26.

2.5-52 Jung, T-P., Makeig, S., Stensmo, M., and Sejnowski, T.J., "Estimating Alertness from the EEG Power Spectrum", IEEE Transactions on Biomedical Engineering, 44, 1, 1997, pp 60-69.

2.5-53 Wilson, G.F., and Eggemeier, T., "Psychophysiological Assessment of Workload in Multi-Task Environments", in Damos, D., (Ed) "Multiple Task Performance", Washington, DC, Taylor and Francis Press, 1991, pp 329-360.

2.5-54 Gevins, A.S., Morgan, N.H., Bressler, S.L., Cutillo, B.A., White, R.M., Illes, J., Greer, D.S., Doyle, J.C., and Zeitlin, G.M., "Human Neuroelectric Patterns Predict Performance Accuracy", Science, 235, 1987, pp 580-585.

2.5-55 Gevins, A.S., Bressler, S.L., Cutillo, B.A., Illes, J., Fowler-White, R.M., Miller, J., Stern, J., Jex, H., "Effects of Prolonged Mental Work on Functional Brain Topography", Electroencephalography and Clinical Neurophysiology, 76, 1990, pp 339-350.

2.5-56 Wilson, G.F., and Fisher, F., "Cognitive Task Classification Based upon Topographic EEG Data", Biological Psychology, 40, 1995, pp 239-250.

2.5-57 McMillan, G.R., Eggleston, R.G., and Anderson, T.R., "Nonconventional Controls", in Salvendy, G., (Ed) "Handbook of Human Factors and Ergonomics, 2nd Edition", New York, NY, Wiley, 1997, pp 729-771.

3. THE INTEGRATION OF ALTERNATIVE CONTROL TECHNOLOGIES

3.1 INTRODUCTION

3.1.1 INTEGRATION AND DESIGN

Before any novel control device can be introduced into a man-machine interface it is necessary to consider how it can be installed and how it should be used. This section therefore discusses the problems of successful integration from two viewpoints:

- A human factors approach which examines the ideal design process and discusses the human factors tools that are available to develop, refine, and evaluate interfaces. This approach is designed to capitalise on the strengths and minimise the weaknesses of human operators.
- An engineering framework which examines mechanical and electrical issues associated with the selection and location of new components in the crew station; and the computational architecture required to interpret nonconventional control outputs from the human and integrate them with the outputs from conventional controls.

In our view, these approaches are complementary and should be pursued in a concurrent fashion. Section 3.2 addresses the ways in which a human factors approach can be integrated into the design process, while section 3.3 focuses on the engineering framework.

All design processes need to integrate requirements from different sources, including commercial, political, practical, managerial, engineering and end-user pressures. For a human-machine interface design, operator requirements should be a major influence. If a decision has already been made to use alternative control technologies (ACTS) for reasons other than human factors (for example there may be workspace limitations which prevent conventional controls being used), there are still many human factors methods which will help integrate the new control technology. Preferably, however, the decision to use ACTs within an interface should be made as the best operator-centred solution within the existing engineering constraints.

Design processes are becoming more multi-disciplinary and concurrent and many design drivers, such as usability and maintainability, can influence a new design at much earlier stages than was previously possible. Ideally, a human-centred design philosophy should be adopted to ensure that a system is designed starting with the operator's (and the maintainer's) interface and using human factors principles. The final design always will be a compromise, but it should be based on an understanding of what the best human factors design is, and how much performance is lost by modifying it in various ways. There will often be some up-front constraints, such as maximum space, or particular lighting conditions. These should be minimised as far as possible if we are to discover what the optimum interface would be. If time and resources allow, designing with complete independence from up-front constraints is a very useful preliminary exercise which creates a human performance yardstick against which to compare subsequent design solutions.

As human-in-the-loop systems grow more complex, the interface between the human and the system is becoming a potential weak link. In some cases, the human's task also becomes more complex, requiring more interaction with the system, and bottleneck problems may occur at the interface. In other cases, the task becomes apparently simpler (e.g. if many tasks become automated) but there may be significant problems in keeping the human sufficiently informed about what is happening in the system. ACTs may be identified as helpful in both such "underload" and "overload" situations. Some might increase the communication bandwidth between the human and the system, thus enabling a greater amount and complexity of information exchange. Others could allow the control of systems to be carried out in more subtle ways than is currently possible, such that certain natural behaviours could be monitored and used as triggers to modify what the system is doing. This would allow very sophisticated systems to be developed. However, if the human factors and engineering issues are not sufficiently addressed, the potential advantages of ACTs may not be delivered.

3.1.2 ACTs AS SUPPLEMENTS AND SUBSTITUTES

An existing interface may be modified to include an alternative control technology. This may be as a supplement, for example if a new task is being introduced; or it may be a substitute for another control, perhaps in order to improve human performance. In both cases, it is important to analyse the whole set of tasks which the operator has to carry out, and not just focus on the local context. The modification of an interface can significantly change the whole task for the operator, and introduce unexpected knock-on effects (for example, a pilot's situation awareness might be affected by making a function less explicit). This means that it is essential to use task analysis to ensure that all inter-relationships have been accounted for, and the designer should also consult human factors databases with respect to the human processing involved.

The analysis of an existing task with an existing interface is simpler than dealing with hypothetical designs, in that it involves factors which can be observed and measured. The effect of introducing a new component, such as an ACT, can thus be measured to a large extent. However, the choice and implementation of the new component should be made with an understanding of the demands that the task places upon the human operator, and the alternative ways in which it is possible for the operator to do the task.

3.1.3 FUTURE INTERFACE DEVELOPMENTS

In current practice, an interface based entirely on human characteristics is impossible to achieve because of the many constraints which inevitably apply, such as cost or technological capability. It is likely that we shall come closest to the ideal interface design with virtual reality interfaces, if future developments in technology and human factors allow this, simply because of the virtual environment's potential low-cost configurability and independence from physical factors. However, the configuration of the interface is not the only consideration. The relative roles of the human and the machine are changing. The machine is now increasingly

capable of sophisticated behaviours, and therefore the contexts and rules of interaction could become quite complex. The relationship between the operator and the machine may not necessarily be that of a master and slave (where command inputs have a fixed structure and meaning). The operator and machine could work more like a team, such as supervisor and operator, where the machine is capable of inference, adaptation to changing circumstances, and complex decision making. The human-plus-machine would become in effect a joint cognitive system. Although potentially capable of powerful and sophisticated performance, such systems may have incomplete information about each other's intentions, very much like human-to-human team work. The interface is thus more dependent upon communication of uncertain and potentially ambiguous information. Cognitive ergonomics has arisen as a human factors discipline to address specifically the integration of the user's psychology and the system's information processing, and it is hoped that progress in this field will help resolve some of these significant problems. ACTs have a very important role to play in making such interfaces possible by broadening the scope of human-machine communication. But it is the interpretation of an ACT output which requires considerable human factors knowledge and engineering development.

3.2 HUMAN FACTORS

3.2.1 THE BENEFITS OF HUMAN FACTORS AND HUMAN-CENTRED INTEGRATION

One of the major benefits of incorporating human factors into the design process is that it brings the discipline of methodical analysis of factors which are too easily taken for granted or overlooked. When humans are in the loop there are so many permutations and unknown factors which can affect the way in which tasks are carried out, that they have to be modelled in a very different way than the rest of the system. Even an experienced design team with a good understanding of the task and the conditions, may not forestall all major problems. Human factors practitioners have methods and models which can help to ensure that certain factors are addressed, and can represent and predict some aspects of human performance in a system. These methods and models have to deal with imprecise, incomplete and uncertain data, yet provide a useful input to the design process. This is done by combining what is known about human physiology and psychology with experimental studies and performance probabilities, and using methods which force designs to make allowance for the range of variance which is probable for a particular population of operators and the particular conditions in which they will carry out the tasks. The use of ACTs requires a greater knowledge of physiology and psychology than conventional control methods. In particular, data are required about natural behaviour (eye movements, body movements, non-verbal communication), response times and sensitivities in different modalities, and sensory-cognitive compatibility. An understanding of sensory integration is also important when multiple modalities are used together in a task. Much information is available in the literature and in human factors databases (see section 3.2.3.5), but this is by no means comprehensive, and more experimental work is needed to generate the required data.

As human factors becomes more mature as a discipline, it is developing a greater capability to prescribe optimum design solutions, rather than analyse and assess pre-specified options. This requires sophisticated analysis and modelling of sensory and cognitive aspects of a task, even before any technological assumptions have come into effect. The capability to specify human factors requirements in this way is essential for a human-centred design of a completely new interface, especially the sort of joint cognitive systems described above (3.1.3). It is this aspect of human factors which is at the forefront of current human factors research, and the availability of ACTs will make it more possible to implement the prescriptions which human factors will be able to make in the future.

The results of incorporating human factors into a design have been sometimes dramatic but more often subtle, and evidenced in a lack of errors and problems. A user-centred design could offer radical solutions by identifying unconventional interface techniques. If these are not yet technologically possible, they have at least provided a valuable aim for technology research and development. The benefits of human factors input should be far reaching in terms of learning to use the system, increasing the capabilities of the human, obtaining reliable performance with the system, recovery from emergencies, and reducing fatigue.

3.2.2 HUMAN FACTORS IN THE DESIGN PROCESS

Currently most human factors design for aerospace-related interfaces is for modifications rather than completely new concepts. The methods used in these circumstances therefore consist, at best, of analysing the effects of a design modification (detailed task analysis and performance modelling) and evaluating them with a sample of users. Often the choice of a new interface technology will simply be based on a comparison of two or more possible solutions which have been mocked-up as a rapid prototype. In any case, the newly modified interface should at least be compared with the previous version of the interface to ensure that performance has not deteriorated in unexpected ways.

If a completely new, human-centred interface is being developed, a comprehensive interface design process could involve the following steps:

- Identify top-level task requirements (e.g. Mission Analysis)
- Analyse and model the task (allocation of function, i.e. what the operator will do and when, what the machine will do, and when; but not how)
- Determine what communication needs to take place between the human and the machine
- Develop recommendations for interaction requirements (dialogue) based on the type of communication and contexts
- Develop initial recommendations for interface technology, based on type of interaction requirements
- Develop initial design specifications for the interface content based on detailed task analysis, human factors guidelines and predictive models of human performance
- Produce a (rapid) prototype of the design
- Evaluate the design with user trials to establish how the operator does perform his allotted functions

- Re-iterate as required to achieve the required human performance.

The last four steps should be part of any interface design process, and as many of the preceding steps as possible should be included.

Human factors tools used in interface design process could include: task analysis and modelling methods, human performance models and databases, guidelines, design philosophies, simulation, experimental investigation, and human performance metrics.

A review of some of the main tools in the context of ACTs is presented in sections 3.2.3.1 to 3.2.3.9.

3.2.3 HUMAN FACTORS TOOLS

3.2.3.1 Design Principles and Frameworks

There are human factors 'principles' which have been compiled to provide high level aims when designing an interface. For example, Schneiderman [3-1] describes eight main principles:

- Dialogues should be consistent
- Systems should allow shortcuts through some parts of familiar dialogue
- Dialogues should offer informative feedback
- Sequences of dialogues should be organised into logical groups
- Systems should offer simple error handling
- Systems should allow actions to be reversed
- Systems should allow experienced users to feel they are in control, rather than that the system is in control
- Systems should aim to reduce short term memory load (users should not be expected to remember much).

Dix et al. [3-2] describes three principles which can be summarised as: Learnability, Flexibility and Robustness. These principles would suggest that ACTs should, among other things, take advantage of natural behaviour (which requires little learning), use general (flexible) interaction rules rather than task-specific rules, and should provide transparent feedback so that any errors are understood.

Other guideline principles include: Williges, Williges and Elkerton [3-3] who offer seven dimensions (Compatibility, Consistency, Memory, Structure, Feedback, Workload and Individualisation); and a substantial set of guidelines (679) can be found in Smith and Mosier [3-4], parts of which are relevant to integrating ACTs.

Design frameworks are useful in providing a structure to the design space, and ensuring that designers consider alternatives, and all issues associated with particular design decisions. A breakdown of dialogue parameters provides a framework for analysing the sort of interaction which is required by a task. This allows the designer to determine whether a particular control method can be implemented to match a particular type of task. There are five aspects to any interaction (see for example [3-5]), including those achieved with alternative control technologies:

- Style (how fixed the interaction meanings are; for example interfaces may be constant or adaptive; [3-6]
- Structure (how constrained the rules are, e.g. protocols or natural language; see for example [3-7])
- Content (how explicit or implicit conveyed information can be, e.g. semantic codes or raw data)
- Context (how context dependent the interaction is; including contexts such as system failure)
- Mode (see for example [3-6]). Traditionally only two modes have been identified: verbal and spatial. This may be insufficient for interfaces with ACTs, where implicit pilot state modes may be used).

McMillan, Eggleston and Anderson [3-8] describe two types of paradigm for coupling operator intentions to machine activation.: the Servo Paradigm and the Structural Coupling Paradigm. The servo paradigm (effectively a monologue rather than a dialogue) involves the operator making pre-determined, intentional command actions to invoke fixed machine responses, while the structural coupling paradigm views the operator as a performer, whose performance is monitored in order to ascertain what the machine should be doing. Many ACTs can be implemented under either paradigm, but the structural coupling paradigm will require ACTs which can monitor the performer. In addition, the structural coupling paradigm requires other sources of information (e.g. about vehicle status) and an inference engine, in order to interpret the appropriate action to instigate in the machine. In this framework, the human 'operator' becomes another variable which has to be taken into account by the system to determine the most appropriate action. The inference engine must therefore be primed with considerable knowledge about the operator. This theme is explored further in section 3.3.

The choice of coupling paradigm will depend upon the particular circumstances into which an ACT is being integrated. The servo paradigm is the simplest, and is likely to be more appropriate for 'upgrading' an existing workstation which already uses servo coupling for other control functions. The structural coupling paradigm is more appropriate as an overall framework for the whole of the control interface.

3.2.3.2 Allocation of Function

It seems obvious that functions should be allocated according to capability (such as human workload capacity or decision making abilities, and machine data processing). If this is the only consideration, however, the overall performance of the human and system may be sub-optimal or problematic.

It may be beneficial to allocate tasks to the operator purely to ensure that one particular event is related to another (e.g. so that the operator is aware of what is happenning). Also there are considerations such as maintaining alertness, job satisfaction, retention of training, error avoidance, and crew interaction, which may have implications for allocation of function.

When bringing ACTs to an interface, allocation of function should refer not only to the sharing of tasks between human and machine, but also to the allocation of tasks to different (sensory) modalities. There is also the possibility that a task might be carried out with different control methods under different circumstances, or even by allowing the operator to choose.

There are several approaches to allocation of function. It is not considered good practice to simply allocate as much as possible to the machine, as this risks leaving the human with unrelated tasks or work underload. There are at least four ways in which functions are allocated [3-9]:

- allocation to machine by a priori management decisions
- allocation according to respective capabilities
- allocation by formal analysis of tasks and sub-tasks
- allocation by Fitts' list [3-10].

The first two methods are appropriate for tasks where there are constraints which cannot be removed (e.g. only the human may make the decision to attack). The last method, using Fitts' list involves looking up a specific function (e.g. data sensing) and reading off lists of pros and cons for human and machine in performing that function. This method has the disadvantage that it does not easily show how one task can be shared between human and machines, and is not therefore useful for the more advanced approaches to interface design which have been described above as 'joint cognitive systems'.

There is no magic formula to prescribe optimal allocation of function, but any approach should be based on a formal analysis of the tasks at different levels of detail at different points in the design process. Methods used to support the design of ACT-based interfaces should have the capability to allocate to different modalities, as well as to human or machine. For more sophisticated interfaces, the method should allow dynamic and adaptive allocation and be suitable for a joint-cognitive systems design.

Meister [3-11] has developed a substantial method for allocation of function, but this does not handle dynamic or adaptive allocation. Other methods worth looking at are described in [3-12 and 3-13]. Allocation of function methods generally do not allow for allocations between operators in a team, so at present are seldom useful for joint cognitive system approaches.

In many cases, allocation of function can only be carried out by comparing several different allocation solutions. The comparison should be made by means of user trials and/or human performance modelling, comparing workload, performance, error etc. One formal method for doing this could use PUMA, which is a computer-based modelling tool ([3-14]; see section 3.2.3.6).

3.2.3.3 Task Analysis and Modelling

A formal representation of the task will allow inter-dependencies to be assessed, problem areas to be identified, and important aspects of the task to be taken into account. It also creates an auditable trail of the factors which contribute to the design process. The formal representation can be based upon an analysis of different factors, such as processes, functions, goals, human knowledge or skills (see section 3.2.3.4 Taxonomies, below). The process of describing tasks in this way can result in a descriptive 'model' of the task (e.g. a model of the human knowledge applied during a task), which can then be used for assessing workload, cognitive demands, error probabilities etc.

The tool used for such modelling consists of a formal language or notation (diagrammatic, algebraic or textual) which can analyse and describe the whole task in the required level of detail. However, even simple notations can become very complex representations of tasks, as each activity can be decomposed into ever finer components. It is therefore essential to identify precisely which factors and what level of detail are required, before starting an analysis. In order to assess the impact of an ACT, analysis should be carried out to the level of detail which shows how the ACT is employed in individual transactions, for all the relevant types of transaction. This will help identify any physical and cognitive resources needed to use the ACT, and avoid the creation of incompatible or over-demanding tasks. It is important that any control action is compatible with surrounding tasks and goals. For example, if a task is being carried out as a series of manual switches, the introduction of one voice-operated switch might disrupt the integrity of the task. This should be identified by grouping tasks which go together. When 'naturalistic' behaviour is being harnessed for control purposes, as is often the case with ACTs, it is important that the task analysis picks out all natural behaviours and factors which affect them (e.g. where the operator is likely to be looking, or what his body language would be, and environmental conditions such as vibration or light levels). This will ensure that bad designs, such as one which requires the operator to look upwards in response to a low pitched tone, will not arise.

In view of the potential for creating unwieldy data, task analysis should be used selectively, following decomposition paths only as far as is cost-effective. STANAG 3994 and US MIL-H-46855B specify a 'critical task analysis' to be carried out on those tasks which are predicted to have a high workload, or which are critical to safety or mission success. This requires tabulation of information for each task covering a number of categories (such as information required, actions, decisions, environment factors).

A flexible method which can be used at different levels of detail is Hierarchical Task Analysis [3-15 and 3-16]. It allows actions and goals to be grouped and linked, so that a designer can ensure that an overall task structure is retained while making local changes such as the implementation of different control technologies. This means it also can be used for ACTs as supplements or substitutes, as well as for new interfaces.

Hierarchical task analysis is often only the first stage in an analysis, which might go on to incorporate models of human processing to predict time, errors, etc. For example Goals, Operators, Methods and Selection Rules (GOMS) [3-17] is a family of simple models which can be applied to basic procedural tasks to provide time information, which in turn can provide an input to Executive Process - Interactive Control (EPIC) ([3-18]; see section 3.2.3.6).

Kirwan and Ainsworth provide a guide to task analysis [3-19]. Two major failings with many task analysis techniques is that they are not very effective at representing parallel activities, and they tend to represent only 'legal' activities (i.e. do not take account of errors, system failures, unexpected events, etc.). There are methods specifically addressed at calculating error probabilities (see Error modelling, section 3.2.3.7).

The data for a task analysis can be time consuming to generate and unwieldy to manipulate. Some techniques can be implemented using computer-based tools, to assist with the data handling. Data flow software, such as MicroSaint (Rapid

Data Ltd) can be used to generate functional decompositions with associated timings. 'Tree' is a useful software tool for assisting with hierarchical task analysis, and is available on the internet (http://www.fineline-software.co.uk).

3.2.3.4 Taxonomies

Taxonomies delineate the classes or categories into which a task or activity can be separated, such as actions, skills, performance, knowledge etc. They are important in achieving consistency and repeatability in analyses, especially with respect to level of detail (i.e. they keep analyses at the same level). In this sense, taxonomies define the units of analysis. At the lower levels, where finer detail is required, task-specificity is inevitable, and it may be necessary to introduce extra categories, or even create a whole new taxonomy. At higher levels, however, it may be possible to refer to more generic taxonomies. Gawron et al. [3-20] conducted a review of taxonomies which cover a range of hierarchical levels, and developed a general human factors taxonomy.

An appropriate taxonomy for integrating ACTs within an interface would define the sorts of function each ACT could do (e.g. track, select, indicate stress). It would be useful if the taxonomy could also capture the features by which ACTs could benefit interaction (e.g. 'track target (eyes)', showing that eyes always look at an object to be tracked; this highlights a possible exploitation of natural behaviour, such as eye pointing). There is thus a need to develop a control taxonomy especially for ACTs.

As control actions are in turn controlled by cognitive activities, consideration of the cognitive basis for control is also required in order to differentiate control actions which are implicit or explicit, automatic or mediated by focused attention, routine or part of a new strategy, etc. These factors will be important in selecting and implementing ACTs, and a taxonomy which includes cognitive factors should be used (e.g. [3-21]).

3.2.3.5 Human Factors Databases

Many databases are being compiled to collate the vast literature which applies to human factors design activities. Much of the information is in the form of experimental results, showing such factors as perception thresholds, sensory facilitation or inhibition effects, and response times. Care must always be taken in identifying the source of such data, as often the source studies are limited laboratory experiments or specific applied cases. Such data can provide guidance, but may not be directly applicable to a given interface problem.

A major human factors database available to the general public can be found on Computer Aided Systems Human Engineering Performance Visualisation System (CASHE PVS), a CD-ROM tool [3-22] including the Engineering Data Compendium (Boff and Lincoln, [3-23]), MIL STD-1472D 'Human Engineering Design Criteria for Military Systems, Equipment and Facilities', and a Perception and Performance Prototyper (which allows the designer to visualise the effects of varying design parameters).

CSERIAC (Crew System Ergonomics Information Analysis Center, Wright Patterson Air Force Base, USA) and EIAC (Ergonomics Information Analysis Centre, University of Birmingham, UK) are major repositories of human factors information and provide services to carry out literature reviews on particular topics.

Databases could encourage a narrow approach to human factors, as they refer to particular components of an interaction. This may lead to a component by component design approach. Such an approach will certainly not guarantee that the whole interface will be optimised. As a well-worn maxim from Gestalt psychology and systems engineering says: the whole is greater than the sum of its parts. However, if a database is used along with the other tools described here, such problems should not arise. There are a number of initiatives which aim to develop a fully integrated human factors design tool, into which the various databases are embedded, and these are discussed in section 3.2.3.6 below.

3.2.3.6 Predictive Modelling

There are many human performance models, created for different purposes and with varying degrees of validation. Some aspects of human performance (e.g. visual performance) are more amenable to validated modelling than others (e.g. cognitive functions). The models are useful as a first pass to evaluate an interface design. By providing the model with data about the proposed design, it is possible to predict some aspects of human performance, such as mental workload, thereby identifying human processing bottlenecks or, occasionally, underload. An expert system, HOPE (Human Operator Performance Evaluator), is being developed by AGARD WG 22 to assist in the selection of human performance models. Some models of relevance to ACTs are listed below, but as always these should be put into the wider context of the whole task performance.

There are several, rather similar, tools for predicting workload with different task and interface designs. Most are based upon Wickens' [3-24] psychological theory of multiple resource pools. They generally involve a task analysis being entered either on the basis of a design proposal or from observation of an existing task, and an assessment of workload over time is generated. This allows the identification of potential overload or underload problems. By changing aspects of the task such as which ACT is used for a specific action, comparisons can be made. Such tools include: POP (Predictor of Operator Performance) from DERA, UK; W/INDEX (Workload Index) from Honeywell, USA; PUMA (Performance and Usability Modelling Tool) from Roke Manor Research, Siemens, UK; WINCREW (Windows version of US Army's workload and crew complement tool CREWCUT) from Micro Analysis and Design, USA. WINCREW and W/INDEX are particularly easy to use, while PUMA is more sophisticated.

Models of human anatomy and biomechanics (anthropometric manikins) such as JACK (Transom Technologies, Inc., USA) and SAMMIE (SAMMI CADL, University of Nottingham, UK) are implemented in a computer-based environment into which geometric information about the workstation can be imported. Various anthropometric analyses can then be carried out, such as testing for reach or muscular force demands. Also different user populations can be specified (e.g. 95th percentile European female).

There are some collections of human performance models and data which attempt to provide an integrated tool for human factors analysis. They tend to have various modules which

are selected and run together in response to inputs about the task. As these are complex tools, expertise is needed to run them and to make use of their output. This is a very ambitious aim and it is likely to be some time, if ever, before a fully comprehensive and validated tool is achieved. In the meantime, it will continue to be necessary to apply some human factors expertise in the use of the various individual tools currently available.

A number of integrated human factors tools are described below.

HOS (Human Operator Simulator) (see NATO Technical report AC/243: A Directory of Human Performance Models for System Design) allows the simulation of a total man-machine system performing a complex mission by integrating human performance into a system modelling framework.

EPIC (Executive Process-Interactive Control) [3-18] is a model of human processing which can be run as a simulation. It incorporates recent theoretical and empirical data, and has separate elements for perception, motor processes, and cognition. This is particularly good for ACTs as it can simulate human performance in multi-modal interfaces. It imports a GOMS analysis (see 3.2.3.3), and predicts human performance and timeline data.

COGNET (Cognition as a Network of Tasks) from CHI Systems Inc. is a highly developed and sophisticated tool for building models of human performance. It is based on cognitive task analysis, but is able to address specifically motor knowledge and control activities. It can provide a total framework of the design of an interface (for example, [3-25])

IPME (Integrated Performance Modelling Environment) from MicroAnalysis and Design, contains a number of human performance models (e.g. anthropometric and physical data, visual performance, mental workload), but is more useful for physical based tasks than complex cognitive tasks.

MIDAS (Man-Machine Integrated Design and Analysis System) from Westinghouse [3-26] is a complex simulation test-bed which in response to inputs of crew station design, operator characteristics and tasks, provides an output of human factors results, task performance and a visualisation of the simulated task.

3.2.3.7 Error Modelling

The assessment of error likelihood is important for any interface, but there are particular considerations for the sorts of interfaces likely to be using ACTs. ACTs by their very nature are designed to make use of 'natural' human behaviour, and therefore error takes on an additional meaning in this context. Natural human behaviour is perhaps more prone than other interface behaviour to contextual influences. It is also perhaps intrinsically more variable. This is one reason why redundancy is important, ensuring that more than one behaviour is used to control an input. So for some ACT-based interfaces, the analysis of error should be broadened to include an analysis of variability in behaviour, perhaps through observation or experimentation if there are no existing previous data. Variability can occur between individuals, but also within an individual over time. There are particular challenges for 'joint cognitive systems' as errors in implementing a control may be caused by errors in the machine interpretation of the human intentions, rather than human error or variability.

It is important to have models of both error occurrence and error recovery. In many cases it may be more efficient to design a system which allows rapid and effective error recovery, than to try and reduce the probability of error to an acceptably low value.

The simplest method of error analysis is to use the sequence of discrete activities performed by the operator, derived from the task analysis, to also generate an 'error tree' in which major errors are broken down into their sub-components, and each sub-component is allocated a probability of occurrence (preferably based on real data). The sum of the sub-component probabilities provides an assessment of the probability of the major error. For example, flying to the wrong waypoint might be a function of a number of sub-errors such as heading selection error, mis-reading of instructions, bad decision making, bad situation awareness. Each of these may be caused by other errors, such as manual/vocal transmission error, wrong expectations, confusion of symbols, etc. Existing databases of human failure probabilities are based on models of human error and data collected from various sources, both in the field and in the laboratory. We do not have a database specifically for ACTs (with the exception of speech), and it would be valuable for any trials with the technology to report errors and variance. Obviously performance is very technology-dependent, and until the technology is reliable some of the more subtle errors may not become evident.

There are also error prediction methods which rely on the judgements of subject-matter experts (i.e. experienced operators) but when introducing novel controls, such experts do not yet exist. If it is possible to do some laboratory simulations to elicit error data, this would be advisable. However, it is unlikely that a single laboratory would have the time and resource to carry out an exhaustive assessment of errors, and so only a rough guide would be obtained.

It may be possible to make some limited use of existing error data for other interfaces. The THERP Handbook [3-27] is a set of data tables providing error probabilities for a range of tasks in the nuclear industry, but some of the data could be applicable to the assessment of error likelihood with ACTs. Also Generic Data Tables [3-28] can provide a rough guide, but context effects are not allowed for.

For ACTs, therefore, it is important to take an approach to error assessment which firstly can identify the potential cause of errors and the context dependency of those causes, and secondly allows task-based probabilities of errors to be mapped onto the causation model. Hollnagel's CREAM [3-29] breaks down error consequences into error 'modes', which in turn are broken down into person related causes and system related causes. These are in turn broken down into malfunctions which can be related to specific psychological error mechanisms, and the factors which influence them.

3.2.3.8 Rapid Prototyping

It is preferable to rapid prototype only after having analysed and designed the interface according to guidelines, otherwise the assessment will be on a hit-and-miss basis. However, if only limited options exist anyway, rapid prototyping can be used to compare those designs, and to assess human performance. But the value of the exercise is dependent on how the evaluations are made. Rapid prototyping makes use of tools to assist in creating an adequate representation of the

interface for evaluation purposes. These tools may be software to generate computer graphical representations with some level of functionality (e.g. VAPS (Virtual Prototypes Inc), Designer's Workbench (Coryphaeus Inc), or Virtual Reality), or simple physical mock-ups with behind-the-scenes human substitutes for machine functionality. (See next section, 3.2.3.9 for a discussion of evaluations).

3.2.3.9 Evaluation and Performance Measures

Evaluation of each integrated ACT is different, and evaluation trials must be tailored to particular requirements, conditions of use, etc. It is important to identify at the outset what criteria are important for the assessment and how they can be measured. Some criteria may include:

- Compatibility (perhaps indirectly measured by ease of learning, error rate, intuitiveness, reduced workload, better situation awareness)
- Capability (perhaps measured by faster response times, greater accuracy, capacity for parallel activities)
- Reliability (perhaps measured by fewer errors, less variability)
- Flexibility (perhaps measured by ease of reconfiguration and reallocation, versatility achieved by operators)
- Acceptability (perhaps measured by user ratings, trials in the workplace, analysis of socio-cultural context).

Response times and errors are relatively easily measured. The measures are meaningless, however, unless they are made within the context of a careful experimental design which takes into account the user sample, the control of variables, the order in which tests are made, the way in which experimental participants are briefed, and the statistical techniques used to evaluate the results.

Sometimes it is not possible to conduct experimental trials, and a small sample of potential users are called in to provide their subjective opinions. Even with such limited control over the evaluation, it is possible to maximise the value of information collected by using carefully constructed questionnaires or interviews, using a 'Think Aloud' method [3-30] or by testing mental workload or situation awareness achieved with the prototype interface.

The measurement of workload and situation awareness would be very valuable in assessing the impact of the whole interface: the sum of its various components. Several tools are being developed to try and do this. There is a difficulty in defining exactly what workload and situation awareness are, and there are many varied opinions on this subject. However, the tools mentioned below can be interpreted as saying something about the 'goodness' of an interface, whether or not the assumptions on which they are based are in fact valid.

NASA TLX (Task Load Index) and variants, is a NASA - developed tool in the form of a paper or computer-based questionnaire, to assess subjective workload.

SWAT (Subjective Workload Assessment Technique), developed by Wright Patterson Air Force Base, is an on-line subjective tool with 3 'domains' of workload, each of which is given a rating between 1 and 3 at critical parts of a task. Recent developments with this technique have enabled the identification of a 'red-line' of maximum workload, above which performance falls off.

SAGAT (Situation Awareness Global Assessment) was developed by Northrop as a direct measure of situation awareness, but as it requires task interruptions, it is of limited use, and may be unacceptable to users in evaluation trials.

SART (Situation Awareness Rating Technique), from DERA is a questionnaire for assessing subjective situation awareness, which has been refined through repeated use. There are several versions, including CC-SART which aims to assess the cognitive compatibility of interfaces.

Other measures worth considering are eye movement patterns, blink rate, heart rate, galvanic skin response, performance on a secondary task (to assess workload), training time, behaviour modification (e.g. effect of new technology on behaviour patterns). It is important to assess the variance of such measures both within individuals over time, and between different individuals.

When the results of evaluations have been gathered, it should be possible to identify significant performance problems or design weaknesses. Consideration should be given to the design of both the task and the interface when seeking to improve performance, and this can be done with the help of task analysis and performance modelling.

It must be emphasised that the design and evaluation process should be regarded as an iterative cycle which starts by putting together a human-machine interface which has been recommended through an analysis of requirements. It involves conducting a series of evaluations, primarily so that the design team can understand its usage and eliminate unforeseen defects before the arrangement is frozen and built. The foregoing material discusses key ergonomic and psychological issues and describes a variety of tools that can materially assist this process. Ultimately the design team must use their judgement in choosing which issues to address and in selecting the most appropriate tools for their application.

3.3 ENGINEERING INTEGRATION

3.3.1 A WORST CASE EXAMPLE: COMBAT AIRCRAFT

The satisfactory introduction of novel controls into any working context raises a gamut of engineering issues, and it is assumed that the procurement of the equipment would be carried out to comply with a wide range of engineering standards specified by the customers. If we take an advanced but otherwise conventionally operated combat aircraft as a reasonably worst case example these standards would, for instance, have been relevant parts of Military Standards [3-31] in the U.S.A. or Defence Standards [3-32] in the U.K., but the increasing emphasis on commercial standards should be noted. Although these documents guide the designers and cover most of the detailed engineering concerns, the introduction of novel controls would necessitate some re-consideration of the physical arrangement of the cockpit systems and the flow of information between the cockpit and the remainder of the aircraft systems.

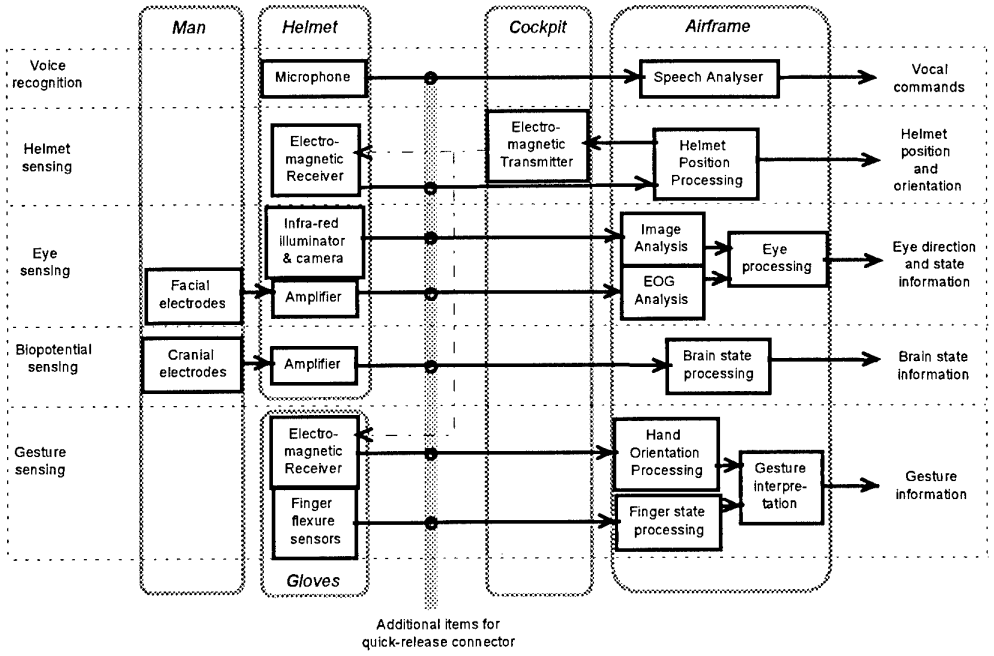


Figure 3-1 A schematic representation of the location of novel control system components.

Mechanical and Electrical Design

3.3.2 MECHANICAL AND ELECTRICAL DESIGN

As shown in the simplified schematic, Figure 3-1, the integration of novel controls could involve mounting components in the airframe, the cockpit, on clothing and on the pilot.

Using current electronic packaging techniques, the airframe-mounted units which carry out signal handling and information processing functions would be electronic boxes. Their form would depend upon the avionic architecture of the aircraft. Although the fitting of additional Air Transport Racking (ATR) boxes is invariably difficult in the full equipment bays of in-service aircraft, the future adoption of modular avionic systems should enable additional facilities to be accommodated relatively easily [3-33]. Such architectures, in which processing is carried out in a set of standard units, are intrinsically reconfigurable and to a high degree fault-tolerant.

If an electro-magnetic helmet and hand tracking system were adopted as part of the alternative control suite, the transmitter would be the only cockpit-mounted component. This would probably be bonded to the inside of the canopy just above the rear of the helmet where it would be close to the receiver but out of the pilot's view and furthest from any metal structure. If, however, conductive coatings are applied to the canopy, this siting may need reconsideration. The adoption of an optical tracking system would, as noted in section 2.2, require different considerations over the location of the sensors to minimise reflections and avoid the capture of direct sunlight.

To avoid interference with existing systems, particularly the ejection seat, the elements which are attached directly to the crewmember or his helmet and clothing, must be arranged to separate automatically from the other cockpit-mounted or airframe-mounted units on ejection. The perennial problem of arranging electrical cabling and connectors would be likely to continue. As indicated in Figure 3-1, the increase in the number of signals and low voltage power lines coupling the cockpit to the aircrew headgear and clothing will further complicate any quickly detachable, automatic mechanism, such as the personal equipment connector (PEC) on the ejection seat. Alternative approaches, for instance using multiplexed fibre-optic channels to transfer data in digital form, become increasingly attractive.

The design of aviator helmets which incorporate night vision sensors and display devices in addition to the original protective, life support and communication functions is already a serious challenge, primarily because this integration must be done without increasing the headborne mass or introducing any compromise which would make the headgear less acceptable in service [3-34]. The technical problem of the "integrated helmet" has stimulated considerable ingenuity among avionics system manufacturers who, collectively, have developed a variety of approaches. Each technical choice offers advantages and disadvantages, and, given the complexity of the trade-off between factors which control the overall result, manufacturers naturally advocate their individual designs. There is, however, no consensus or acknowledged preference for a particular approach. When extra elements, such as the eye image sensor and an array of active electrodes, are included in the list of requirements, the difficulty of producing satisfactory headgear will certainly be exaggerated. The topic will persist, with developers pursuing and exploiting every relevant technological improvement as it becomes practical.

3.3.3 COMPUTATIONAL DESIGN

Having arranged the satisfactory physical installation of the novel control suite, it is necessary to connect it to the rest of the avionics. In this example application, the aircraft is likely to include intelligent advisory aids and some pilot state

monitoring [3-34] in addition to comprehensive mission, utility, flight control and weapon systems. It is suggested that the novel control suite would be integrated with the conventional controls, as shown schematically in Fig. 3-2.

The additional function, called a "command interpreter" adjudicates between the signals generated by any of the novel or conventional control modalities in order to send an unambiguous command to the relevant aircraft system. Such a command interpretation function is already incorporated in advanced fighters, for instance Rafale and Eurofighter, primarily as a means of integrating the voice control system in these aircraft.

The fundamental requirement is that the command interpreter should accept only the intended (Intention being defined by the pilot, by doctrine, by tactics, and by other factors.) switch selections, utterances, designations, gestures and perhaps mental states of the pilot. Unintended utterances, designations, gestures and mental states should be identified and have no effect on the aircraft systems. This could be accomplished by software in which the allowable links between the received output from the controls and the signals which are sent to the systems could be specified as finite-state "rules". These would trap errors, such as double selections, and express the constraints and flexibilities which, by analogy with man-computer interaction, constitute the aircraft's operating system. It is these rules which would embody the statement of what is desirable from consideration of the human factors discussed in section 3.2. For instance, the ability to select an external object by fixating with the eye, pointing the helmet sight or indicating by a hand-pointing gesture, then saying "target", "lock radar" or "range", or the flexibility to perform a mixture of these actions, would be programmed as a set of command acceptance rules.

Although the embodied rules would be unlikely, for instance, to allow weapon release by voice command, the software must be treated as crucial to the safety of the vehicle. It would therefore be produced to avionic standards and be coded in an approved high-level language, such as ADA, and be subjected to rigorous verification and validation testing.

The command interpreter could produce three classes of output in addition to the main collated system control signals. As indicated in Fig. 3-2, these would be:

(a) Information fed back to the display system so that the user can be kept aware of the state of the control mechanisms in order to operate them satisfactorily. For instance, this would include the highlighting of a virtual key selected by eye fixation, the visual and auditory presentation of the output from the speech recognition system, and the movement of a cursor responding to finger pointing.

(b) Synergistic feedback to the novel control suite, to enhance the performance of one sensing systems using information derived from other sensing systems. For instance, speech recognition reliability could benefit from knowledge of eye-pointing direction by biasing the context to the objects or selections near the pilot's fixation.

(c) The novel controls produce additional pilot state information, for instance the pilot's eye pupil diameter and his blink rate from the eye tracker, his formant frequency from the speech recogniser and his head activity from the helmet tracker. All of these would supply extra information which could assist the intelligent aid which monitors the pilot's state to make a more reliable classification, for instance whether workload had induced boredom or frenzy, and whether he was conscious.

Although the need for co-ordination between displays and controls, covered by (a), would be assumed by system integrators, the rôle of the command interpreter as a means of producing the extra benefits of synergistic feedback (b) and additional pilot state information (c) may be surprising. It is thus evident that the organisation and structure of the software which carries out the command interpretation function is central to successful exploitation of the novel controls. In essence, although the behaviour of individual controls would depend largely on the characteristics of the individual modalities, the programming of the command interpreter would dictate the usability of the set.

It is of interest to consider how the "control metamorphosis" espoused by McMillan, Eggleston, and Anderson [3-8] could be implemented in a future aircraft. In this postulated evolution, the introduction of non-contact control mechanisms allows a progression from the conventional one-way "command", via a two-way "dialogue" to a "structural

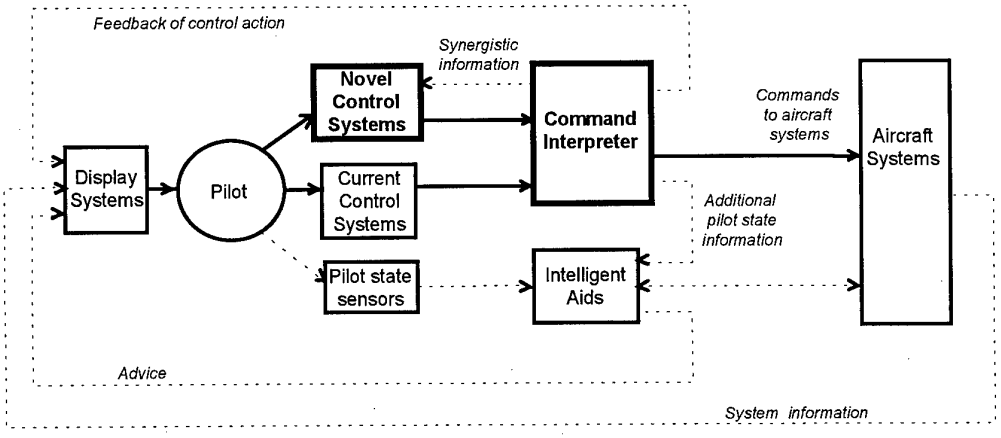


Figure 3-2 Broad information flow paths in an aircraft with novel control systems.

coupling" paradigm with a "transactional" interactive style. In the last approach the dumb controls are replaced by an intelligent "agent" which senses operator outputs, infers operator state and exerts direct control over sub-systems in a manner which is consistent with the operator's intent. The pilot could then cease to operate the aircraft on the basis of a master-slave relationship and, instead the pilot and agent would mutually adjust to each other to achieve an implicit shared goal. In terms of the structure shown in fig. 3-2, the agent would consist of a suite of suitably powerful and autonomous intelligent aids together with the command interpreter. The existing path between command interpreter and the aircraft systems would be severed, and all control would be mediated by the agent.

3.3.4 ADDITIONAL FACILITIES: CONTROL OF THE CONTROLS

Conventional control mechanisms all have fixed characteristics and cannot be matched to the qualities of the operator. In contrast, novel mechanisms are themselves likely to require "control controls" so that they can be set up to suit the individual user and be de-selected in the event of failure. This is certainly a peculiar entailment.

The most evident need is to be able to calibrate the appropriate sensing system to match the possible mixture of voice, eye, hand gesture and cortical response characteristics of the user in order to optimise accuracy and reliability. Ideally any time-consuming calibration, such as fixating in sequence on a target array or speaking a list of commands, should be possible in calm circumstances. It is suggested that any pre-flight procedures should be simple, quick checks to confirm correct system function and perhaps set an alignment, and it should not be necessary to engineer facilities which allow re-calibration in flight. Any lengthy procedures should occur on the ground, for instance in the training simulator. A comprehensive simulation facility which represents the whole man-machine interface will, in any case, be vital to enable pilots to explore and gain confidence in the use of the alternative controls.

As well as a means of gaining confidence and familiarity, the pilot could be given the freedom to use the training environment to explore the merits and penalties of alternative ways of setting the controls and displays. He could then tailor the man-machine interface to suit his preferences, in a manner analogous to a computer user setting the relationship between mouse and cursor or the disposition of opened windows on the screen. In this case it may be that different pilots would prefer different command words, or need a different repertoire of gestures. Such "preference" information would be transferred with the calibration information from the training environment to the operational aircraft using pre-flight data link or portable operational data storage facilities. Although a "default" set-up could be available and imposed if it was considered necessary, it is highly unlikely that the freedom to optimise the interface would be treated capriciously. All pilots would be most concerned to evolve a system which they could remember how to use, and which enabled them to effect control accurately and reliably.

Finally, the high level means of exercising control over the novel controls could probably be engineered by something as simple and unambiguous as a dedicated panel housing a short row of "on/off" toggle switches. These would be marked clearly to identy each modality. When the user becomes convinced that for instance his spoken commands are interpreted erroneously, he can switch the voice recognition system off knowing that the command interpreter will be aware of this de-activation, and he can continue the mission using the remaining facilities. However, it must be admitted that this issue warrants further study, particularly to understand how quickly the user is likely to become aware of control system mal-functions, and whether the command interpreter could also undertake this monitoring function.

3.3.5 REFERENCES

3-1 Schneiderman, B., "Designing the User Interface: Strategies for Effective Human-Computer Interaction", Massachusetts, Addison Wesley, 1987.

3-2 Dix, A., Finlay, J., Aboud, G. And Beale, R., "Human-Computer Interaction", Hemel Hempstead, UK: Prentice Hall, 1993.

3-3 Williges, R.C., Williges, B. And Elkerton, J., "Software Interface Design", in Salvendy, G., (Ed) "Handbook of Human Factors", New York, NY, Wiley, 1987.

3-4 Smith, S.L. and Mosier, J.N., "Design guidelines for user-system interface software", The Mitre Corporation Technical Report ESD-TR-84-358, 1984.

3-5 Barnard, P.J., and Hammond, N.V., "Cognitive contexts and interactive communication", IBM Hursley Human Factors Laboratory Report, 1983.

3-6 Elkerton, J. And Williges, R.C., "Dialog design for intelligent interfaces", in Hancock, P.A., and Chignell, M.H. (Eds) "Intelligent Interfaces: Theory, Research and Design", Amsterdam, Elsevier, 1989.

3-7 Moran, T.P., "The Command Language Grammar: A representation for the user interface of interactive computer systems", International Journal of Man-Machine Studies, 15, 1981, pp. 3-50.

3-8 McMillan, G.R., Eggleston, R.G., and Anderson, T.R., "Nonconventional Controls", in Salvendy, G. (Ed), "Handbook of Human factors and Ergonomics, 2nd Edition", New York, NY, Wiley, 1997.

3-9 Bailey, B., "Human Performance Engineering", Englewood Cliffs, NJ, Prentice Hall, 1982.

3-10 Bekey, G.A., "The human operator in control systems", in De Greene, K.B. (Ed) "Systems Psychology", New York, NY, McGraw Hill, 1970.

3-11 Meister, D., "Behavioural analysis and measurement methods", New York, NY, Wiley, 1985.

3-12 Shoval, S., Koren, Y., and Borenstein, J., "Optimal task allocation in task agent control space", in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 4, 1993, pp. 27-32.

3-13 Rencken, W.D., and Durrant-Whyte, H.F., "A quantitative model for adaptive task allocation in human-computer interfaces", in IEEE Transactions

on Systems, Man and Cybernetics, 23 (4), 1993, pp. 1072-1090.

3-14 Warren, C.P., Day, P.O., Hook, M.K. and Hicks, M., "Performance and Usability Modelling in Air traffic control (PUMA)", in Proceedings of the Fourth International Conference on Human-Machine Interaction and Artificial Intelligence in Aerospace, Toulouse, Sept. 28-30, 1993.

3-15 Annett, J., Duncan, K.D., Stammers, R.B. and Gray, M.J., "Task Analysis", Training information Paper no. 6., London., HMSO, 1971.

3-16 Shepherd, A., "Hierarchical task Analysis and Training Decisions", Programmed Learning and Educational Technology, 22, pp. 162-176, 1985.

3-17 Kieras, D.E., Towards a practical GOMS model methodology for user interface design", in Helander, M. (Ed) "Handbook of Human-Computer Interaction", Amsterdam, North Holland Elsevier, 1988, pp. 135-158.

3-18 Kieras, D.E., Wood, S.D., and Meyer, D.E., "Predictive Engineering Using the EPIC Architecture for a High Performance Task", in Proceedings of CHI'95, ACM, 1995.

3-19 Kirwan, B., and Ainsworth, L.K., "A Guide to Task Analysis", London, Taylor & Francis, 1992.

3-20 Gawron, V.J., Anno, G., Fleishman, E.A., Jones, E.D., Lovesey, E.J., McGlynn, L.E., McMillan, G., McNally, R.E., Meister, D., O'Brien, L., Promisel, D.M., Ramirez, T., "Human Factors Taxonomy", in Proceedings of the Human Factors 35th Annual Meeting, 1991, pp. 1282-1287.

3-21 Rasmussen, J., Pejtersen, A.M. and Schmidt, K., "Taxonomy for Cognitive Work Analysis", in Proceedings of the First MOHAWK Workshop, Liege, May 15-16, 1990, Vol. 1 pp. 3-153.

3-22 CASHE:PVS, Version 1.0, Computer Aided Systems Human Engineering Performance Visualisation System. CSERIAC, Products Department, AL/CFH/CSERIAC Bldg 248, 2255 H Street, WPAFB OH 45433-7022, USA. (Email: cseriac@falcon.al.wpafb.af.mil).

3-23 Boff, K.R., and Lincoln, J.E., "Engineering Data Compendium, Human Perception and Performance, Vols. I, II, and III", Armstrong Aerospace Medical Research laboratory, Ohio, 1988.

3-24 C.D. Wickens, "The structure of processing resources" in Nickerson, R. And Pew, R. (Eds) "Attention and Performance VIII", Hillsdale, NJ, Erlbaum.

3-25 Ryder, J. And Zachary, W., "Experimental validation of attention switching component of the COGNET framework", in Proceedings of the Human Factors Society 35th Annual Meeting, 1991.

3-26 Corker, K.M. and Smith, B.R., "An Architecture and model for cognitive engineering simulation analysis : Application to advanced aviation automation", in Proceedings of the 9th AIAA conference on Computing in Aerospace, New York. 1993, pp. 1079 - 1088.

3-27 Swain, A.D., and Guttman, G., "Handbook for Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications", Report NUREG/CR-1278, US Nuclear Regulatory Commission, Washington, DC, 1983.

3-28 Kirwan, B., "A Guide to Practical Human Reliability assessment", London, Taylor and Francis, 1994.

3-29 Hollnagel, E., and Cacciabue, P.C., "Cognitive Modelling in System Simulation", in Proceedings of the Third European Conference on Cognitive Science Approaches to Process Control, 2-6 September, Cardiff, UK, 1991.

3-30 van Someren, M.W., Barnard, Y.F., and Sandberg, J.A.C., "The Think Aloud Method, A Practical Guide to Modelling Cognitive Processes", London, Academic Press, 1994.

3-31 MIL-STD-1776A (USAF) Aircrew Station and Passenger Accommodations, 1994

3-32 Def Stan "Design and Airworthiness Requirements for Service Aircraft Vol 1- Aeroplanes" AVP-970 published by MOD(PE) London (This is an evolving document in which Chapter 107 covers "Pilot's Cockpit - Controls and Instruments" and Chapter 105 covers "Crew Stations - General Requirements"

3-33 Jarrett D. N Karavis A " Integrated flying helmets" Proc Instn Mech Engrs Vol 206, pp47-61 1992

3-34 "Safety Network to Detect Performance Degradation and Pilot Incapacitation" Papers presented at AMP Symposium. Tours, France. April 1990 AGARD-CP-490

4. SOME PROPOSED APPLICATIONS OF ALTERNATIVE CONTROLS

4.1 INTRODUCTION

In Chapter 1 there were a number of areas described that would be suitable for alternative control techniques, and the proposition that use of those techniques would bring operational benefits. Of the five alternative control systems, head tracking and speech control are considered the most mature, with head tracking control already being in operational use. Eye tracking, whilst technologically advanced and useable in simulators, etc., still requires some development for use in the environment of the airborne cockpit. Gesture-based and biopotential-based systems are regarded as longer term solutions for the airborne cockpit, although, of course, many simple EMG systems are regularly in use as aids for persons with disabilities.

There is a gradual transition to use of some of the alternative controls and this is apparent in the next generation operational aircraft of the Eurofighter, F-22 and Rafale type. Experimental flying in the UK, France and USA are using helmet mounted sights and displays which are fully dependent upon head tracking to provide the helmet pointing capability. Some production systems are in service use in a number of countries (Russia, Romania, South Africa, Israel, etc.) and are generally simple sights, but, even at this stage of development, allow significant increases in operational performance, when correctly integrated with a suitable weapon system. Similarly voice control systems are flying experimentally and being used during simulation to demonstrate significant operational benefits. Eurofighter will use voice control as an integrated part of avionics control and Rafale will have the use as an option. Eye tracking would appear to have strong potential in military systems, particularly when used in conjunction with head tracking. Whilst head tracking gives a good indication of the position in which the head is pointing - and is fully adequate for a number of applications - it does not, of course, necessarily show what the pilot is actually looking at (i.e. where the gaze is fixed). A number of techniques exist to measure eye position, some more robust than others in an operational environment, and many are used in the simulation environment, where many practical environmental limitations are minimised. The electrophysiological measures, and measures of gesture, are currently limited in their application to the operational cockpit, although there is considerable potential for these type of systems in the 2010 to 2015 timescales. The adaptive interface between the pilot and aircraft systems, in which the aircraft will infer the state of the pilot and the pilot and aircraft will have a knowledge of the state of each others systems, will need the capabilities of these technologies.

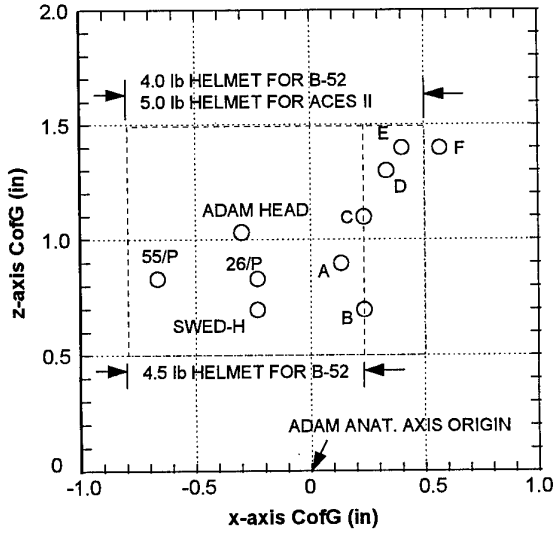


Figure 4-1 Interim Ejection Criteria

A further issue that must be addressed arises from the implications that the alternative control technologies involved with head pointing, eye tracking and voice control all require some sort of head mounted system. In the cockpit environment these systems are generally head mounted on the flying helmet or an existing helmet mounted display. However, care must be taken to minimise both the total head mounted mass and the centre of gravity (CofG) offsets caused by the additional head mounted equipment, as there are flight safety issues involving injury to aircrew from such systems, which will have serious implications in operational use. Current helmets fitted with HMDs weigh in the region of 1.8 to 2.2 kg and generally the centre-of-gravity of the helmet is high in the 'z' (vertical) axis, due mainly to the optical train being located high on the helmet, and forward in the 'x' axis. During high G manoeuvres, or during ejection, the turning moments on the neck, due to the combination of the offset CofG and helmet mass, can be high enough to cause injury - serious or otherwise - to the neck and spine. To minimise the risk of injury the CofG of the complete helmet system should be kept as close to the natural centre of gravity of the head as possible, and the mass of the helmet ideally, at least for head mobility and comfort issues, not higher than 1.4 to 1.6 kg. The current (1997) targets for helmet mass are 1.4kg (3.1 lb.) for the UK and 4.4 lb. for the US Air Force and 4.0 lb. for the US Navy. The US requirement also notes that 'lighter helmets (3.5 to 4.0 lb.) *are* recommended to enhance overall aviators' performance'. An example of one of the USAF helmet/head centre of gravity limits for ejection are shown in the interim ejection criteria in Figure 4-1, which is for helmets of given mass, 4 and 5 lb. in this case, and for class of ejection seat.

For the helicopter cases, the highest risk is in the crash case and there are two criteria that have been discussed. One is based upon the Mertz criterion and involves procedures developed from vehicle crash research - which is predominantly in the horizontal and lateral axes, and some work in France validating the model in the vertical axis. The criterion is based upon the dynamic stresses imposed upon the neck in an axial direction, gained from experimentation with anthropomorphic dummies, and involves a comparison of acceptable loads on the neck against a time history. Figure 4-2 shows the acceptability criterion. This procedure takes into account the dynamics of a crash, but cannot be generally used for predictive estimates of risk, although a family of curves could be built up from the model.

The US Army, which is again predominantly for helicopter operations, uses the USAARL curves, one of which is shown in Figure 4-3, which uses a family of curves plotting head supported mass against CofG in the 'z' and 'x' axes. These graphs are used predominantly for the helicopter crash case, as high G manoeuvres and ejection cases are inappropriate, at this time at least, for helicopters. The criteria are based on a series of static

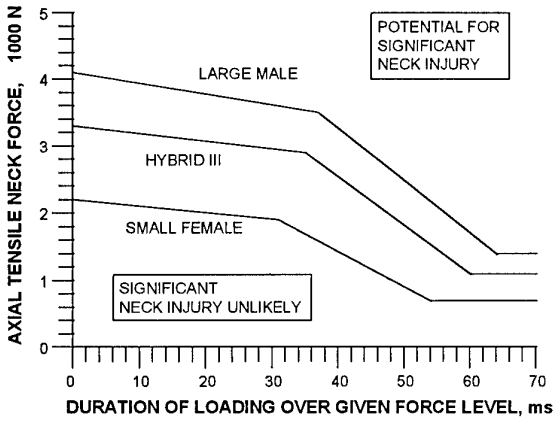


Figure 4-2 Acceptable loads on the neck

calculations and plots the mass of the helmet against the centre of gravity in the appropriate axis, with acceptability criteria.

The two methods approach the problem from differing directions and some differences are apparent, particularly between static and dynamic analysis. There is a general agreement that comfort is related more to CofG location, either in the 'x' (fore and aft) or 'z' (vertical) axis, but the two methods differ in the effect of CofG on injury risk. The dynamic approach [4-1, 4-2] suggests a lesser sensitivity of injury risk to CofG location, but a significant effect of helmet mass. The USAARL approach [4-3] has, at this stage, a more even weighting between helmet mass and CofG, based on a curve of constant mass moment, and the authors recognise that validation is required using dynamic simulation techniques.

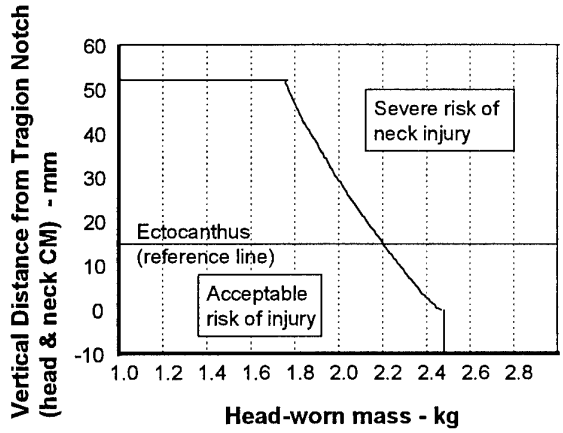


Figure 4-3 Head supported mass vs. Centre of Gravity

The applications discussed in this document have been specifically constrained to those for aviation application. Many of the technologies described, however, are more generic in their use and there are numerous additional applications outside of the aviation scenarios. Experimental work is under way in areas of tele-medicine, robotics etc., and over the next decade use of the technologies described in these pages will allow multi-modal dialogues with a wide range of systems to become common place and the interactions between the operator and system to be more natural and less complex.

4.2 COCKPIT/CREW STATION APPLICATIONS

4.2.1 HEAD MOVEMENT AND POSITION TRACKING

Of the potential alternative control technologies, head movement tracking is undoubtedly the most used and is, and has been, in operational use in a number of forms, for a number of years, with a number of airforces. Such tracking technology is of primary importance as it is fundamental to all future systems that will wish to use head or helmet mounted displays - in any conceivable form - or any type of visually coupled system.

There are a number of technologies available to provide head tracking, the most mature being the AC magnetic tracker. This system, in its well developed form, provides the bulk of the systems in use in the world. It provides good accuracy when mapped into individual cockpits, and, as importantly, it provides a large headbox in the cockpit allowing the pilot to obtain head position data from essentially the whole effective cockpit volume. Accuracy is variable across the cockpit volume, particularly towards the extremities of the headbox, but accuracies in the region of 4mR rms., with no worse than 8 mR are achievable (1998), and degradation of the sensor signals are generally regarded as 'graceful'. In the first instance, this is fully acceptable for use in close combat in 'one v one' or 'one v more' engagements with short range missiles. The missile seeker head is slaved to the pilot's head movement and the field of regard of the seeker head is large enough to negate the need for very high accuracy (< 3 mR). Significant savings in target engagement times are achievable with this simple system. For air-to-ground targeting, this simple system also has some advantages. With a targeting FLIR, the FOV is generally small, typically 10 degrees, and searching for a ground target with a hand controller and head down display is difficult and time consuming. Designation of the ground target through the helmet sight, with the FLIR slaved to the head movement, will again allow significant time savings to be made in target designation, particularly for targets of opportunity. As part of this targeting system, it would be feasible to use the voice command as a control medium, - an alternative to using a manual action - to mark and lock onto the ground target. Where 'eyes out' operations are at a premium and the manual processes require glances inside the cockpit, voice command would provide an alternative control strategy.

In a similar way, the use of visually coupled systems, to allow a pilot to fly operationally in night and poor weather conditions by the use of sensor derived images, place an essential need upon the measurement of head position. In helicopters of the US Apache type, the Pilot Night Vision System (PNVS) couples the nose mounted infra-red sensor directly to the pilot's eye-piece on the helmet display, and slews the sensor to the pilot's head position, lining up the sensor point of regard with the pilot's eye, using the optical head tracking system. Similar systems exist on a number of other helicopters, either experimentally or close to operational deployment. In current operational deployment, the pilot retains a view of the outside world to help retain some external situational awareness, but eventually it may be necessary to operate without a direct external view, if the ultimate protection against optical blinding weapons (e.g. lasers) and other considerations become necessary. This situation then, essentially, becomes the beginning

of the "virtual cockpit" and some of the limitations of the systems needed to enable implementation of this philosophy become apparent. One of the major limitations is in the temporal delays (also called transport delays or latencies) that are induced by the system processing needed between the sensor inputs from the outside world and the sensor output to the crew's eyes. For instance, the delays between the pilot moving his head to look at a particular point in space and the sensor fixing on that same point should be minimised, and, whilst not yet well defined by flight experimentation, figures of 20 to 50 ms are often quoted - the shorter the better. The delays depend upon the level of processing needed in both the head movement and the image transfer system, and the more complex the system, generally the longer the delays. For next generation helmet mounted displays, where, for instance, flat panel displays may be the image source on the helmet, the considerable levels of processing needed to carry out, for instance, distortion correction in off-axis optical systems, needs good and careful system design to minimise transport delays.

One of the assumptions with head tracking is that the head position reflects accurately where the eyes are looking. Whilst this may be broadly true, there are, of course, many times when this will not be the case. Measurement of eye position and gaze point, along with head position, will nullify the errors in these assumptions.

4.2.2 EYE TRACKING

There are a number of distinct advantages to using eye position sensing as an operational tool. A primary advantage is the increased surface area that can be swept by the eye in comparison with the head movement range in the cockpit environment. Figure 1-6 in Chapter 1 shows this clearly, with the average horizontal range increasing in the order of 50 degrees and the vertical range by 35 degrees. This increase in the swept spherical surface area should provide the ability to locate and designate a wider range of targets particularly in air combat, and this may be of a particular advantage during high G manoeuvres, where head movement may be restricted, but eye movement is not, and an adequate field for targeting available, from eye movement alone, with the head being necessarily held in a fixed position during the high G manoeuvering part of flight.

The introduction of large screen displays in surveillance and C^3I type aircraft (and land vehicles and sea control rooms) may also provide some advantage for eye position tracking. In a complex and detailed screen picture, particularly where the eye needs to scan the full picture for information, the position of the cursor is not always immediately apparent and time is lost in search. By designating the required spot, say a target designator (TD) box, by fixing the eye upon it and calling for the cursor to that position by voice command, considerable time savings should be apparent. Similarly, in such tactical displays updating the TD boxes should be accomplished more quickly by the use of eye designation of the particular box and voice command to 'switch', than by conventional manual means in many cases. Such control systems would be used as an alternative method to the manual process and such multi-modal dialogues would allow considerably greater flexibility into the control of complex display systems.

A further application, in air-to-ground attack, could be improvements in the accuracy of designation, by the eye, of ground targets for smart weapons, the accuracy being improved by the capability of the eye to be naturally physiologically stabilised in the presence of turbulence or the low frequency buffet normally present in the ground attack phase. The accuracy of head movement measurement under these conditions of turbulence is unlikely to be of adequate quality for precision designation [4-4].

One spin-off application of viewing the eye in an eye tracking system is the potential capability of monitoring the eye state (pupil size, blink rate, etc.) as part of the aircraft adaptive interface that may be feasible in providing adaptive controls, displays and aircraft systems that respond, in some way, to the changing levels of aircrew state.

4.2.3 VOICE RECOGNITION OR DIRECT VOICE INPUT (DVI)

Although not quite as mature as head tracking, there have been significant increases in capability in the last few years and this technology now has the potential to enter service with operational aircraft in short timescales. The technology is, at least in speaker dependent systems, at a level that will allow it to compete successfully against manual switching for a number of operational tasks. A number of experimental flight trials have been carried out, in a number of countries, and the results, in terms of technical capability, are essentially identical across countries in the normal cockpit environment. However, the extremes of the operational envelope, i.e. under high G, reduce the recognition rates of DVI systems. The ability to fly 'eyes-out' whilst switching systems that normally require a visual operation within the cockpit, can provide a significant advantage in the heat of operations and enable reductions in crew workload. Experimental flying has shown DVI clearly as an enabling technology, particularly in the areas of complex or time consuming switching. As an example, radio channel selection or in the areas of switching the modes of map displays or other displays, DVI has an important role to play and it is often the additional time-consuming detail of systems operation, which causes interference with sequential operational tasks, that causes high workload for aircrew. The use of DVI where the feedback of the successful recognition is clearly visible or audible is clearly more appropriate for aircraft use at this stage of potential operational use. However, even where the feedback loop is not directly perceived, DVI can be used successfully and trials in ground simulators during helicopter tactical operations have shown that digit strings, for instance waypoint entries into navigation computers, can be easily accomplished with audible feedback of the recognised digits being fed to the ear in real time. This is particularly so in the cases of aircrew who regularly listen and communicate across a number of radio nets. Thus feedback of the digit, word or phrase that the system has recognised can be either overt in terms of audio or visual feedback or covert in terms of the pilot recognising that the voice command has actually changed the display or system (e.g. a map display changing scales from 50,000 to 250,000).

The acceptance by aircrew of such voice command systems has been generally high in all countries where experimental flying has taken place. In 1993 RAE trials in the UK in a Tornado aircraft, 65% of aircrew thought that the system was acceptable, with small improvements needed, 25% rated the system as neutral (i.e. neither good nor bad) and the remaining 10% thought it unacceptable and needing major improvements. None, however, thought it totally unacceptable. Since those trials there have been improvements in the technology of recognition systems and acceptability ratings should have improved.

Similarly in the EF2000 programme the use of DVI in the active cockpit simulator has shown the operational effectiveness of

using such systems, and the systems are fully supported by pilot opinion

The extremes of the cockpit environment, that is extremes that affect either the speech signal to noise ratio (e.g. noise) or the speech production process itself (e.g. noise, pressure breathing, high G, etc.) still cause some problems with speech recognisers and this is more evident in high-performance fixed wing aircraft than rotary wing aircraft. Thus, at this stage, DVI systems may be more suitable for immediate operational use in helicopters and transport aircraft. For fixed wing applications experimental research has shown that under increasing levels of g_z, relative speech levels increase by up to 13 to 14 dB at $8g_z$, with a large spread, and this, combined with speech spectrum changes, results in a declining recognition performance from around 94 to 96% accuracy at the standard $1g_z$ to around 90% at $5g_z$ The necessity to use pressure breathing at the higher G levels also results in reductions in recognition accuracy, which, with both G and pressure breathing, falls to around 78% at 5g, and 65% at 6g. Thus comparing pressure-breathing and non pressure-breathing conditions at 5g, a loss of some 12% can be attributed to pressure-breathing effects alone.

However, in spite of these performance losses at the extremes of the flight envelope, DVI systems work well in the high percentage of mission time that the operational aircraft spends below 5 to 6 g - and in a well loaded ground-attack aircraft of the Tornado or Harrier type - this is essentially the majority of the time.

The further development of speaker independent systems will reduce the need for training speech recognition systems and further techniques to improve recognition rates when operated under high environmental and battle stress, will, in the near future, make voice recognition a highly flexible alternative control technology.

4.2.4 BIOPOTENTIAL- AND GESTURE-BASED CONTROL

Control based on gesture or biopotentials is probably more futuristic, certainly in the context of the conventional cockpit. However, as the cockpit evolves from the conventional to the virtual cockpit, more opportunities arise. Also the trend towards the use of UAVs gives rise to the potential of both airborne based and ground based 'cockpits' where full or partial control of the UAVs are based on 'man-in-the-loop' principles. As 'cockpits' move away from the harsh environment of existing military aircraft operations, to the potentially more benign conditions of the ground based, or the control-aircraft (civil/transport type) based cockpit, the conditions for use of most alternative control technologies become more attractive, and this is particularly so for gesture and biopotential controls. Environments in aircraft of the C^3I type, AWACS, Rivet Joint, etc., have control areas that will be similar to some airborne based cockpits and perhaps are potentially the first type of aircraft that will be able to make use of these advanced alternative controls.

4.3 CONTROL OF WEARABLE COMPUTERS

With the continuing miniaturisation of head-mounted displays and computer hardware, the potential applications of wearable computer systems are generating a great deal of interest in commercial and military sectors. In the near future, wearable computers are envisioned as key support devices for personnel engaged in maintenance and repair tasks, for emergency medical personnel operating far from their home base, for package delivery personnel, and even to support the navigation, communication and tactical awareness of individual foot soldiers operating in combat environments. Each of these environments generates high manual workload for the user. Ideally, the user's hands should remain dedicated to their primary tasks while other modalities are used to interact with the wearable information system. At the present time, speech recognition is available as a hands-free control technology for wearable computers. However, environmental noise, requirements for simultaneous communication among team members, and the need for quiet, covert operations suggest that other hands-free controllers need to be developed as well. In Section 4.3.1 we will review some of the issues that must be addressed when developing alternative controls for wearable systems. In Section 4.3.2 we will provide a brief overview of the alternative control modalities that are being evaluated for US Air Force maintenance personnel.

4.3.1 SOME ISSUES FOR THE APPLICATION OF ALTERNATIVE CONTROLS TO WEARABLE COMPUTERS

4.3.1.1 Mobility

Just as wearable computers are designed for mobile operators, input devices must also be mobile. Lightweight, compact, and comfortable components are required and wireless transmitter technology should be employed to prevent snagging or entanglement of cables and connectors. Ideally, the input device should be operable regardless of operator movement or position, and easy to don and doff.

4.3.1.2 Task Environment

Hands-free controllers need to operate in any environment in which wearable computers are employed. Environmental factors that may impact the application of hands-free controllers include ambient noise, light, temperature levels, smoke, and industrial contaminants. Some controllers may not be appropriate when privacy or covertness are of concern. Likewise, collaborative use of computing and networking technologies can impact input device choice. Since wearable computers are likely to be employed in the presence of others, the input device and operator responses need to be designed so that there is no interference with ongoing communications and no negative social responses.

4.3.1.3 User Population

Implementation of hands-free controllers will be impacted by whether the general population is targeted or whether the controller can be customised for an individual operator. If the former, the design should assume that the operator has little computer sophistication; procedures need to be obvious, natural, and require little, if any, training. Likewise, the need for operator calibration and adjustments should be minimised or automated. However, for more specialised military or technical applications it may be acceptable, and even advantageous, to customise the controller and/or utilise a longer training protocol. For example, tailoring signal detection algorithms to an operator's EMG response characteristics or training on a speaker dependent speech recognition system may produce significant long-term payoffs.

4.3.1.4 Task Requirements

To date, the control achieved with most hands-free devices can be described as rudimentary. Any application must take into account the limited dimensionality, accuracy, speed, and bandwidth of control afforded by these devices. For some applications and target users (for instance, those with severe physical limitations), speed and accuracy of control may be of less concern, since conventional control options are not possible. Other applications require more rapid and error-free performance. Since hands-free controls do not require associated limb or hand movement, control inputs are typically very rapid, unless lengthy signal processing is required. Rather, it is the precision and accuracy limits of the fundamental human responses and of the controller hardware and software that constrain application. In light of these limitations, efficient procedures are needed for correcting erroneous entries and for safeguarding the system from hazardous control inputs. The characteristics of the task must also be considered. If the content of the data input is not known in advance, then the input device needs to support arbitrary input. In this case keyboard surrogates or speech recognizers are likely candidate devices. If the input content is fairly constrained, and user interaction can be reduced to a selection process, then a more limited input device is acceptable. Concurrent tasking must also be considered; if the operator's visual attention is totally occupied by a task, then the use of gaze pointing for control is not appropriate.

4.3.1.5 Frequency of Control Input

Although wearable computers are essentially "on" all the time, the frequency of operator input can range from sporadic to constant and can vary with task demands. Controller selection should take into consideration the anticipated input frequency. For example, just as extended manual keyboard entry can lead to carpal tunnel syndrome, frequent use of a jaw clench to activate EMG-based control can aggravate TMJ (tempero-mandibular joint) disorders.

4.3.1.6 Controller Activation

Input devices for wearable computers need to be constantly operational or capable of being engaged in minimal time. However, given that many alternative control modalities, e.g., eye gaze, gesture, EMG, and EEG, exhibit natural variation that can appear to be a control input, spurious activations are of concern. One means to address this is to use a multimodal interface that allows one modality to serve as an activation command or consent for another input modality and thus minimise inadvertent inputs.

4.3.2 ALTERNATIVE CONTROLS FOR MAINTENANCE WEARABLE COMPUTERS

As part of an ongoing program in integrated logistics and maintenance, the US Air Force is evaluating a variety of electronic job aids for maintenance personnel. Many types of computer-based systems, including wearable components, are being investigated in this research (Figure 4-4).

In addition to reducing the need to publish and maintain massive numbers of paper documents, these computer-based systems promise: (1) more rapid transmission of current technical data to the field, (2) better integration of logistics and maintenance data bases, (3) automation of many reporting requirements, (4) further standardisation of maintenance procedures, and (5) improved equipment turnaround times and sortie generation rates.

Wearable computers that provide real-time access to

Figure 4-4 Maintenance Technician with Wearable Computer System

technical data and procedures represent one such support system. In early trials, the noise in and around the flight line was found to be a severe challenge for the speech recognition systems provided with the wearable prototypes [4-5]. Accordingly, research efforts are being directed toward improving the performance of flight-line speech recognition and toward the development of other hands-free control options. Because of the highly proceduralized nature of many maintenance tasks, even simple discrete inputs can enable the maintainer to interact with the computer support system. Control modalities under evaluation include the following (the enabling technologies for each of these candidates are described in Chapter 2):

- A variety of techniques to enhance the performance of speech recognizers, including ultrasonic and optical sensing of lip motion and the use of facial EMG signals which are highly correlated with specific word formation patterns.
- Facial gestures, measured via forehead EMG patterns, as discrete inputs. With this approach, specific gestures would be employed to enable and disable control inputs, to step from one procedure to the next, and to confirm the completion of required maintenance items. Preliminary work suggests that such patterns can be detected with electrodes mounted in the headband of a typical head-mounted display.
- Teeth clicks as discrete inputs. Single and double clicks, analogous to mouse button presses, or more complex temporal codes could be employed to perform the functions listed above.
- Eye tracking as a direct pointing input. Although currently available systems do not meet all of the requirements of this application, eye tracking in combination with any of the discrete input modalities could provide a highly flexible means to interact with the maintenance support system.
- EEG as a direct pointing input. Discriminable cortical evoked responses can be produced by simple manipulations such as intensity modulation of icon-sized areas of a display. Implemented in this fashion, a form of EEG-based eye tracking is provided. In combination with the discrete input modalities, this approach could also provide a flexible means of control. Preliminary work suggests that such patterns also can be detected with headband-mounted electrodes.

While the maintenance support software could be designed so that all interactions could be performed with any one of the above modalities, such an approach is likely to be cumbersome and poorly accepted by users. Rather, it is likely that a combination of these modalities will be more desirable, and that conventional controls, such as one-handed keyboards, will be advantageous for activities such as form preparation. It is also likely that the optimal modality will be situation dependent. For example, the user may prefer to employ speech recognition during periods of relative quiet, and turn to another modality when background noise or communication requirements would degrade recognizer performance.

4.3.3 REFERENCES

4-1 Mertz, H.J, "Anthropomorphic Test Devices" in Nahum, A.M. & Melvin, J.M. (eds) "Accidental Injury: Biomechanics and Prevention" Springer-Verlag, New York, 1993

4-2 Lèger, A., Portier, L., Badou, J., and Trosseille, X. "Crash Survivability and Operational Comfort Issues of Helmet Mounted Displays in Helicopters: Simulation Approach and Flight Test Results." Communication Seminar on Helmet Mounted Design, Framingham, Ma. December 1997.

4-3 McEntire, B.J., Shanahan, D.F. "Mass Requirements for Helicopter Aircrew Helmets." AGARD CP597. 1996

4-4 Tatham, N.O. The effect of turbulence on Helmet Mounted sight aiming accuracies AGARD CPP-267 High Speed Low Level Flight: Aircrew Factors 1979

4-5 Chapman, D.D., and Simmons, J.R., "A Comparative Evaluation of Voice Versus Keypad Input For Manipulating Electronic Technical Data For Flight Line Maintenance Technicians", M.S. Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA, Sept. 1995.

5. CHALLENGES AND RECOMMENDATIONS

5.1. FURTHER DEVELOPMENT OF THE ENABLING TECHNOLOGIES

The alternative control technologies reviewed in this report provide promising new methods for interacting with future aerospace systems. Nevertheless, a great deal of research and development is required to achieve the full potential of these new controllers. Some of the these challenges are discussed below.

5.1.1. SPEECH-BASED CONTROL

Continued research is required to improve robustness to new speakers, new dialects, and channel or microphone characteristics. Algorithms that enable automatic speech recognition (ASR) systems to be more robust in dynamic noise environments such as airports or automobiles have been developed, but their performance is still lacking. Speech recognition performance for very large vocabularies and large perplexities is not adequate for applications in any environment. Continued research to improve out-of-vocabulary word rejection, in addition to the above mentioned areas, will enable larger vocabulary ASR systems to be viable for applications in the future.

One answer to the problem of having to remember a large vocabulary is to make the system capable of understanding any command, however it is phrased. The user can then speak naturally, using whatever form of words comes to mind. This removes the workload associated with having to remember which words are valid. Such systems are often called "speech understanding" systems. A reliable natural language interface may be some way off, but is a prime goal for research in speech recognition.

Integration of speech-based and other alternative controls has recently been attempted, but it requires further technology development. The foremost requirement to advance the field of combined lip and speech reading is the collection of a high-quality labelled database. This is needed both for recognizer training and for comparing the accuracy of different methods. Next, exploration of sensory integration and recognition methods that are most appropriate for the unique requirements of speech reading, such as the different time scales between the audio and visual channels and the need for rate invariance, are required.

One key issue that must be addressed is the ability to operate speech-based controls in multi-task environments. Some research has investigated the effect of task loading and other physical stressors on speech and its resultant impact on speech recognition performance. Continued research is needed to reduce the impact of these factors.

5.1.2. HEAD- AND EYE-BASED CONTROL

Magnetic systems are a reasonably mature head tracking technology and are suitable for use in operational aerospace environments. Future challenges include: improving the handling of interference conditions such as moving metal and electromagnetic radiation; reducing calibration and alignment requirements; and improving performance parameters, especially temporal bandwidth.

The combination of magnetic head tracking with other head tracking techniques offers some potential advantages balanced, of course, by the drawback of added complexity. For example, inertial components can be used to significantly increase the temporal bandwidth of the measurements. A simple optical system may provide high accuracy about just the aircraft boresight region (for delivery of boresighted weapons) thus reducing the need for pre-flight alignment of the magnetic system.

Important challenges for eye-based control technology can be divided into equipment development and control protocol development categories. Eye tracking equipment is not yet mature enough for operational flight applications. Corneal reflection/pupil tracking, although not the most precise technique available, provides sufficient accuracy and precision, and comes the closest to meeting the overall needs of military aviation environments. Very significant improvements are needed, however, to make these systems robust enough, dependable enough, and automatic enough for operational use.

Although improvements in accuracy, precision, and temporal bandwidth are all possible, there are inevitable trade-offs between these parameters and robustness in an operational environment. It may sometimes be appropriate to tailor the design of eye control functions to relaxed accuracy and precision requirements in favour of improved robustness, dependability, and ease of use.

Development of effective eye-based strategies and protocols must take account of natural human gaze behaviour. It is important to remember that conscious direction of gaze for control activation is not a natural task. Additional work is warranted on the effective use of visual feedback, on the most effective techniques for confirmation (designation of a particular gaze event to be a control input), on the most effective combinations of gaze with other control modalities, and on the development of guidelines for the characteristics of visual targets used in gaze control tasks.

5.1.3. GESTURE-BASED CONTROL

One of the main drawbacks of most gesture capturing technologies is the fact that they impair movement. Typically a mobile part has to be grabbed, or wires and other obtrusive equipment have to be affixed to the limbs. As a result, the potential benefits of using natural human gestures are severely impaired. Progress must definitely be made in the miniaturisation of sensing devices and autonomous power supplies.

No-contact technologies, such as video-based systems, do not suffer from this problem but do not yet provide adequate performance. Resolution constraints limit the useful range and/or impair the recognition of small features such as the fingers. In addition, video offers limited time resolution, preventing accurate recognition of rapid movements.

Dynamic gesture recognition is not yet a mature field, the main problem being segmentation of the incoming data into high-level gestures. New techniques for detecting the beginning and ending points of a gesture, such as hand

tension and speed, are being considered. The immersion problem, which can lead to the interpretation of gestures that were not intended for the system, can be alleviated by the definition of an active zone, but this is not necessarily appropriate for all applications.

Current interface design seldom capitalises on the richness offered by gesture-based interaction. Much current work is devoted to the development of new, better-adapted interface concepts. Feedback, which has revealed itself as a very helpful feature for precise manipulation, as well as for acceptance of a gestural interface, is still designed in a primitive way. Automatic derivation of adequate feedback stimuli from the nature of the controlled task and its components is needed. Finally, teaching gestural interfaces is still mostly done by example. Adequate notations for gestures are being developed, but there is no consensus to date.

5.1.4. BIOPOTENTIAL-BASED CONTROL

As a stand-alone controller, biopotential-based control is not sufficiently developed to replace traditional control methodologies for aerospace applications. It does, however, hold the promise of providing an intuitive future control alternative and, with creative interface design, almost any discrete response task can be performed using EEG or EMG control. Pattern recognition approaches offer the greatest potential for discriminating signal from artefact and for controlling multiple degree-of-freedom systems. In certain cases, response time advantages have been demonstrated by using EMG-based control for simple discrete tasks. The implementation of proportional EMG control for continuous control tasks is also possible. Continuous EEG control, on the other hand, has not been shown.

With the exception of prosthetic device control, little work had been done outside of the laboratory. There is a profound need to identify applications that require hands-free operation and to develop biopotential controllers for field evaluation. The use of biopotential-based control in combination with traditional or other alternative control technologies should also be investigated. The combined information from the controllers should provide additional discriminating information to reduce command ambiguities and improve system performance.

Biopotential-based control may be most applicable when used in a manner that bridges operator state monitoring and explicit control. Referred to as an intelligent control paradigm, this approach employs an intelligent interpreter that monitors a range of human outputs, including EMG and EEG signals, to infer user intent, a desire for information, etc. The interpreter then issues commands to the system consistent with the inferred user intentions.

Rather than simple substitution of EMG and EEG signals for conventional inputs, optimum utilisation of this technology will result when biopotential information is integrated with information from other sensors to enhance current and future control modalities.

5.2. THE FUTURE OF HUMAN-MACHINE INTEGRATION

As introduced in the "foreword" to this report, the overall aim of those seeking to improve the interaction between an operator and a computer-controlled machine is to sense a wider range of human behaviour, and then use the flexibility which this brings to free the human from having to make more detailed control actions than required by the situation. Although all human behaviour is to some extent learned, the endeavour is to allow the human to point, glance, speak, shout, gesticulate and use other familiar means of communication, rather than have to learn to juggle with a complex array of switches.

In the aviation context the ideal is to give the pilot this human-centred ability to interact with the aircraft systems, but the designers must take account of considerable constraints, most notably that the cockpit is intrinsically inhospitable, the pilot is encumbered by layers of protective equipment and suffers the motion-induced body forces, and he is very concerned with looking out of the window and steering the vehicle. Although Alternative Control Technologies (ACT) will be introduced initially as substitutes for conventional manual controls, it must be emphasised that the extension from this "servo" paradigm to the "structural coupling" necessary to implement a human oriented interface involves profound changes and concomitant uncertainties.

5.2.1. FROM THE HUMAN VIEWPOINT

Designers have access to a fairly extensive knowledge base, and a selection of human factors tools, which can assist them to integrate an ACT into an interface as a substitute or supplement. However, aside from some work on multi-modal dialogue design, there is little to assist them to achieve successful structural coupling. If the servo interface is dumb, in the sense of only accepting a narrow set of instructions, it is reasonable to think of the structural coupling interface as having a co-operative personality in that it actively seeks to help the user. At present it is only possible to imagine some of the complications which could arise.

Essentially the interface must make an active inference as to what the pilot intends to happen, and the "command interpreter" which receives the output from all the control modalities would carry out the sequence of (sense, interpret, decide then effect]) operations analogous to a human operator. The main concern would be to minimise the errors which could occur along this chain and avoid human falibilities. An archtypical example would be the false inference which lead to the killing in 1170 of Thomas Becket in Canterbury cathedral by knights who allegedly misinterpreted King John's exasperation as a fatal command.

Considerable work will be needed to characterise errors, know how to trap them, and give the operator feedback which enables him to supervise the interface easily. Familiarity, training and the establishment of trust will be paramount, and because the accuracy of the systems which provide the sensed data is likely to vary between individuals, it is quite likely that the man-machine relationship could take on human love-hate dimensions. The engineer should seek to have the user regard the co-operative machine as trustworthy, careful, reliable and timely rather than awkward and stupid or eager but naive.

5.2.2. FROM THE MACHINE VIEWPOINT

The principal technical problem for the successful integration of Alternative Control Technologies in the cockpit is the physical difficulty of mounting more devices on the pilot, and ensuring that these to not conflict with the function of other man-mounted equipment or compromise the pilot's safety. The problem reaches a head, literally, at the headgear which is already burdened with the need to satisfy a wide range of protective, life support and display functions while remaining tolerably lightweight.

The problem the system designers face in trying to implement the "structural coupling" paradigm would perhaps benefit from a different conceptual approach. Rather than look at the interface as a set of display and control facilities from the pilot's viewpoint, they must consider also the interface's view of the pilot, in order that they can endow it with the capacity to be able to look at the pilot and include his wishes in a broader view. It is evident that work in this area and in fields such as advanced automation, pilot state monitoring, intelligent aids, joint cognitive systems, human-machine symbiosis, advanced avionic architectures and fault-tolerant computing is all aimed ultimately at a similar objective. Although these areas will retain their particular techniques and are unlikely to coalesce in the near future, it can be hoped that research will benefit greatly from the transfer of ideas and co-ordination between programmes.

5.3. POTENTIAL BENEFITS AND CHALLENGES

5.3.1. BENEFITS AND UNCERTAINTIES

In order to paraphrase a discussion about the way novel control technologies can be exploited, and indicate the remaining challenges facing the system developers, three tables have been constructed.

Table 5-I is an attempt to summarise the detailed usage of control devices, with particular emphasis on their function in combat aircraft. Existing and novel control technologies are listed as the headings for each row, and the classes of current and possible functions performed by control devices are marked on the column headings. The suitability of a particular control device for each class of usage is shown by a solid square marker, while potential uses are shown by question marks. It should be noted that the classes of use are fairly arbitrary and that other classifications are possible, especially when considering man-machine interfaces other than those arranged for military pilots. The control levers operated by the left hand, the right hand and the legs are described in the tables in terms appropriate to a fixed wing aircraft ("throttle", "stick" and "pedals" respectively), but their use also applies to rotary wing aircraft having "collective" and "cyclic" manual levers. The main purpose of Table 5-1 is to indicate the functions which novel devices can perform in place of unsatisfactory manual control devices and to enable new forms of interaction, such as the "virtual cockpit", "pilot state monitoring" and "multi-modal dialogue".

Tables 5-II and 5-III complement the first table by summarising the reasons why some existing devices are unsatisfactory by considering, explicitly, which of the user's needs they fail to meet. These tables also summarise whether novel control devices can be regarded as having overcome the deficiencies of conventional systems. It should be noted that the critical user criteria have been chosen to be most pertinent to the case of combat aircraft, but they are by no means exhaustive and they are unlikely to be regarded by all users as equally important. As with the first table, they are also unlikely to represent the concerns of operators in other contexts, for which alternative relevant criteria would need to be formulated. Where there is reasonable uncertainty that a particular control technology either satisfies or fails to satisfy a particular criterion, this is shown by a question mark. Topics which require further human factors study are marked with an "H", and those which depend upon the technological approach or need further development are shown with a "T", although it can be noted that most need a mixture of human factors experimentation and technological improvement.

The meaning of most of the critical criteria in Tables 5-II and 5-III is clear, but some criteria may need explanation. It should be noted that the first table of the pair deals mainly with the required characteristics of the control system and how well the systems function in the cockpit environment. The second considers aspects which depend on the way the user interacts with the control system, and these are sub-divided into "physical" and "cognitive" criteria.

A control system is "fast" if it gives an output within about 20 msec of an input, and as noted on the first column of Table 5-II, mechanical switches, and head and eye tracking systems are likely to meet this criterion, but speech, gestures and perhaps biopotential sensing systems are not. This is as much a consequence of the time taken to make an input as for the system to recognise it and generate an output. On the other hand, the "short delay between intention and control input" criterion at the end of Table 5-III is meant to reflect the time taken for the operator, having decided for instance to use a guarded switch, to reach for, remove the guarding mechanism, grasp and operate the switch. Comparison between the two columns indicates that most control devices rate similarly on these two criteria and differences arise from the time needed to reach for mechanical devices.

A control is "reliable" in Table 5-II if it makes a consistent response to an operator input, and Figure 5-1, which gives a simplified representation of the information processing stages together with the varieties of feedback information and the relevant errors, may help to explain the difference between this and the criterion of "error resistant", which is the third column of Table 5-III. The important difference is that a control is "reliable" if it gives a consistently correct output signal in response to the action of the operator, irrespective of whether the operator's action is itself correct, and error (D) in Figure 5-1 must be highly improbable. Thus the designers of a speech recogniser, which performs by delivering what is essentially a series of "best guesses", seek to minimise the chance of any misrecognition, and they are very concerned that the few errors which do occur should be easy for the user to spot and correct. On the other hand, the designers of a mechanical switch, which can take on only a few states, aim to make the device so reliable that the possibility of it giving an incorrect output does not need to be considered.

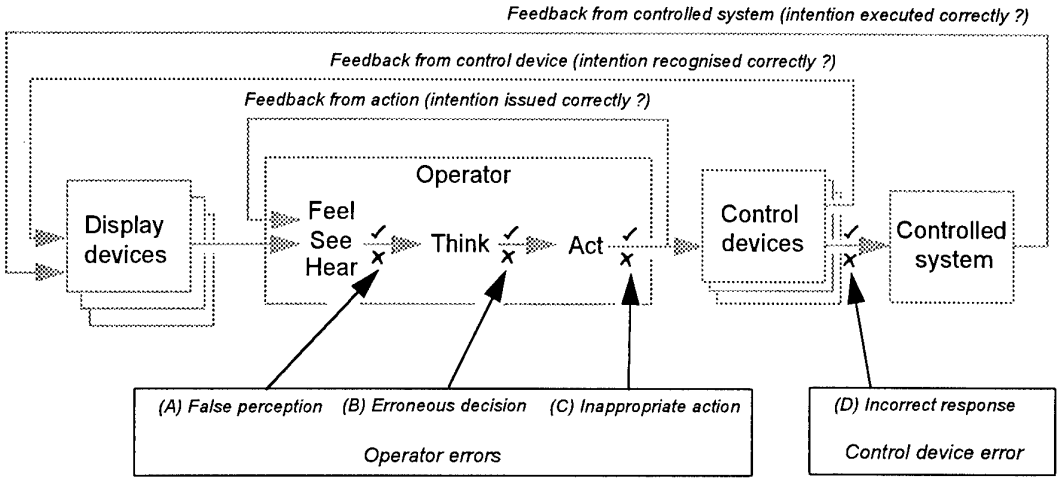


Figure 5-1 A simplified representation of the information processing stages, the varieties of feedback and the varieties of error in a man-machine interface.

The other varieties of error, called "false perception", "erroneous decisions" and "inappropriate actions" in Figure 5-1, for which the operator would undoubtedly be considered culpable, may more justly be attributed to the interaction between the operator and the control device. For instance, mis-recognition of an utterance in a speech recogniser could activate an inappropriate jump within the branching syntax which could lead the operator to take the wrong corrective action. It is also possible that the small movement and force change, which give the kinesthetic and tactile sensations of a key being triggered, could be too delicate and induce the operator to make multiple presses of the key. It is the minimising of such occurrences that is referred to in the third column of Table 5-III as classifying the control device as being "error resistant". Although the fault would be manifested as an inappropriate action by the user, the perceptual and cognitive stages are included in Figure 5-1 because they emphasise that the performance of a control device depends fundamentally on the familiarity of the user with the control device, and on the manner in which the device is integrated into the whole man-machine interface (MMI), particularly on the form of the feedback.

The figure indicates that feedback occurs at several stages along the sequence of processes; firstly on input to the control device, secondly to show the output from the control device, and thirdly giving the response of the system under control. Here mechanical, manual devices score well because they can be arranged to give the natural tactile and kinesthetic sensations, for instance of successfully committing the action of operating a switch or lever. Similarly voice, head and eye based controls would be satisfactory because the user monitors his own speech, even in a noisy cockpit, and he is fundamentally aware of where his eye and head are pointing. However input to the other novel control devices is relatively mysterious; the user would probably not know whether he had made an erroneous gesture or produced an inappropriate level of myoelectric stimulation, and he would certainly have no intrinsic feedback about his electro-encephalographic activity. The first indication he receives in these cases is from the output of the control device, and this explicitly contrived feedback on a visual or aural display is, to some degree, attention-demanding, confusable with other displayed information, and delayed. The delay can induce dynamic peculiarities much like the "head lag" experienced by users of a visually-coupled system incorporating a head tracker. Perhaps the worst aspect is that the user, lacking independent evidence that his actions have been correct, cannot distinguish between his mistakes and those committed by the control device, and is less able to organise an effective correction. This error correction difficulty is indicated in the fifth column in Table 5-III, and should be distinguished from the same criticism applied to manual switches which is a logical consequence of the discrete function performed by a switch. Once a switch signal has been sent it is impossible to retract, although the consequences could range from setting the map to the wrong scale through to the unintended release of a weapon.

5.3.2. CHALLENGES

Perusal of Tables 5-II and 5-III indicates that guarded switches, such as those used to jettison underwing stores in an emergency and to release weapons, will continue to be used to avoid the possibility of inadvertent operation. Similarly, the conventional throttle, stick, pedals, and the switches and inceptors mounted in the grip-tops are largely satisfactory, except perhaps that the latter must be arranged to be operable when wearing layered gloves which tend to hamper dexterity and de-sensitise tactile feedback. Detailed improvement will continue, and it is reasonable to conclude that the provision of gripped inceptors, containing embedded switches operated by the fingers and thumbs, is likely to persist as the fundamental means of steering the aircraft and performing rapid selections. The main challenge is to develop grip-tops and embedded switches which can be operated by a broader range of hand sizes, and avoid increasing the number of such switches in response to the need for confirmation of an input from a novel control device.

The other existing manually operated controls, as described in Chapter 1, are in some way problematical in a combat cockpit because they remove one hand from the primary controls, they are affected by vibration and high-G, and they invariably require the pilots to glance down to locate the

switch so that the hand can be guided to reach it. Keys, particularly "soft keys" whose function depends upon the mode of an adjoining head-down display, and other hand-operated devices such as touch-sensing faceplates on a head-down display, and small joysticks and rollerballs, are more suited to use in the rear of large aircraft where the environment is relatively benign and head-out attention is unnecessary.

As implied by the question marks which populate Tables 5-II and 5-III, utilising all of the novel control technologies involves some uncertainty. Speech recognition systems have been developed to give a satisfactory level of performance in military cockpits and provide an alternative to both dedicated and multi-function switches for carrying out non-spatial interactions such as making selections and data entry. The main challenge is to develop the recogniser to have sufficient accuracy, vocabulary and insensitivity to operating conditions that speech can be used as the preferred and primary means of performing these functions.

Helmet position and orientation sensing systems are an adequately mature technology. Accurate head pointing is humanly practical, and it gives considerable benefit over manual techniques for designating external targets and features. The challenge is one of refining the technologies to give the sub-milliradian accuracy and low latency so that a helmet-mounted display can be used in a visually-coupled system to replace a head-up display in a combat aircraft. Although eye direction measurement is not a mature technology in a combat aircraft, continued technological development is warranted to exploit the likely superiority of human eye pointing over head pointing. The challenge is to integrate the sensing system into the already complex headgear and measure eye direction imperturbably, accurately and rapidly.

Although gesture sensing is a relatively immature set of technologies which are likely to have limited usefulness in combat aircraft, the challenge is to develop recognition systems which are sufficiently robust and accurate that the naturalness of gesture can be exploited as an intrinsic component of a multi-modal dialogue, albeit in less physically demanding environments. Biopotential sensing systems are the least mature class of technologies which could find application in the combat aircraft. Here there are two challenges. The first, which is most daunting, is to show that in the cockpit environment the sensed information can be related to the operator's intentions with sufficient reliability to have the authority to exert direct control over airborne systems. The second, which is perhaps more feasible, is to develop biopotential sensing systems which can be used in conjunction with the other novel control technologies as a means of recognising the operator's physiological and psychological state to create an intelligent interface which actively seeks to assist the user by inferring his intentions.

It is considered that the integration of individual novel control technologies into an aircraft cockpit requires the resolution of a wide range of human factors and engineering issues. Particular care must be taken with the detailed design of the control task, device calibration, the form of the feedback, and the facilities for error correction, and here some guidance is available from relevant human factors experimentation and engineering experience. The main challenge will be to take this one step further; to integrate a group of novel control technologies, together with with compatible manual control devices and advanced display systems, to exploit the complementary functional characteristics of the individual systems and create advanced interfaces giving flexible natural man-machine interaction. Here, a wider range of human factors and engineering issues must be faced, and little human factors guidance is available.

Means of control			Uses: Current — Vehicle steering *eg flight-path control*	Sensor steering *eg RADAR pointing & VCS slewing*	Rare emergency actions *eg jettison stores*	Time-critical selections *eg transmit radio message*	Routine selections *eg change map scale*	Routine data entry *eg insert waypoint co-ordinates*	Cursor slewing *eg move cursor on HDD*	External feature designation *eg designate visible target*	Uses: Future — Virtual switch selection	Pilot state monitoring	Multi-modal dialogue
Existing		Throttle, stick & pedals	■									?	?
		Guarded switch (panel-mounted)			■								
		Grip-top switch				■							?
		Grip-top inceptor		■					■	■	?		?
		Small joystick		■					■		?		?
		Rollerball		■					■		?		?
		Hard keys (single function)				■	■	■					
		Soft keys (multi-function)					■	■					
Novel		Speech sensing			?		■	■			?	?	?
		Head sensing		■					?	■	?	?	?
		Eye sensing		?			?		?	?	?	?	?
		Gesture sensing	?	?			?		?	?	?	?	?
		Biopotential sensing	?		?		?					?	?

■ Accepted ? Possible

Table 5-I Summary of the established and possible uses for both existing and novel means of control in military aircraft

Means of control			Critical environmental and systems criteria								
			Fast (responds within 20 msec)	Reliable (consistently correct response to operator input)	Allows eyes-out operation	High-g tolerant (can be used at +6g)	Vibration tolerant (can be used at 0.25g rms)	Keeps hands on throttle & stick	No setting up needed	Easily used with aircrew protective clothing	Compatible with occluding HMD
	Existing	Throttle, stick & pedals	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Guarded switch (panel-mounted)	✓	✓	✗	✗	✗	✗	✓	✗	✗
		Grip-top switch	✓	✓	✓	✓	✓	✓	✓	✗	✓
		Grip-top inceptor	✓	✓	✓	✓	✓	✓	✓	✗	✓
		Small joystick	✓	✓	✓	H?	H?	✗	✓	✗	H? T?
		Rollerball	✓	✓	H?	H?	H?	✗	✓	✗	H? T?
		Hard keys (single function)	✓	✓	✗	✗	✗	✗	✓	✗	✗
		Soft keys (multi-function)	✓	✓	✗	✗	✗	✗	✓	✗	✗
	Novel	Speech sensing	✗	✓ T?	✓	H? T?	H? T?	✓	T?	✓	✓
		Head sensing	✓	✓	✓	✗	✗	✓	T?	✓	✓
		Eye sensing	✓	✓ T?	✓	H?	✓	✓	✗	✓	✓
		Gesture sensing	✗	H? T?	✓	✗	✗	✗	✗	✗	✓
		Biopotential sensing	T?	T?	✓	H? T?	✓	✓	✗	✓	✓

✓ Satisfactory

H? Questionable - needs further human factors investigation

T? Questionable - depends upon technical approach or needs further technological development

✗ Unsatisfactory

Table 5-II Compliance of control modes with critical environmental and systems criteria in military aircraft

Means of control		Critical operator usage criteria								
		Cognitive					Physical			
		Does not need concentrated attention	Gives confidence that command has been issued, and acted upon	Error resistant (unlikely to induce operator error)	Easy to remember what to do	Allows easy error correction	Suited to multi-modal interactions	Suitable for all likely users	Allows wide range of interactions	Short delay between intention and control input
Existing	Throttle, stick & pedals	✓	✓	✓	✓	✓	✓	✓	✗	✓
	Guarded switch (panel-mounted)	✗	✓	✓	✓	✗	✗	✓	✗	✗
	Grip-top switch	✓	✓	✓	✓	✗	H? T?	✓	✗	✓
	Grip-top inceptor	✓	✓	✓	✓	✓	H? T?	✓	✗	✓
	Small joystick	✓	✓	✓	✓	✓	H? T?	✓	✗	✓
	Rollerball	✓	✓	✓	✓	✓	H? T?	✓	✗	✓
	Hard keys (single function)	✗	✓	✓	✓	✗	✗	✓	✗	✗
	Soft keys (multi-function)	✗	✓	H?	✗	✗	✗	✓	✓	✗
Novel	Speech sensing	✓	H? T?	H? T?	✓	H? T?	✓	H? T?	✓	✗
	Head sensing	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Eye sensing	✓	H? T?	H? T?	✓	H? T?	✓	H? T?	✓	✓
	Gesture sensing	✓	H? T?	H? T?	✓	✗	✓	✓	✓	✗
	Biopotential sensing	T?	H? T?	H? T?	T?	✗	✓	H? T?	H? T?	T?

✓ Satisfactory

H? Questionable - needs further human factors investigation

T? Questionable - depends upon technical approach or needs further technological development

✗ Unsatisfactory

Table 5-III Compliance of control modes with critical operator usage criteria in military aircraft

6. CONCLUSIONS

From a purely human-centred point of view, the use of various control channels appears as totally natural. The sensorimotor system of most living species exhibits a remarkable redundancy, both in term of sensors and effectors (homogeneous) but also by its ability to convey redundant or complementary signals pertaining to the same object via different channels (heterogeneous). Also, neural plasticity and vicariance mechanisms allow development of a large variety of strategies adapted to the situational context. When dealing with « intelligent » agents, this redundancy operates naturally and considerably enhances the dialogue. The generalization of a fully computerized flight system introduces the physical possibility of changing from purely manual controls to the use of various « natural » alternatives to input orders to the aircraft systems. By itself, this opportunity may not be enough, since the robustness and efficacy of manual controls have been widely demonstrated from the beginning of Man-Vehicle history. The first question to answer is therefore about the operational rationale to introduce the required technologies in the cockpit or crew station.

Very demanding operations in the current generations of fixed and rotary wing aircraft considerably increase the need for « eyes out » operation. Meanwhile the complexity of aircraft systems and the speed of operation require that the pilot interact constantly with the system in order to configure it properly and to acquire information. The HOTAS concept, introduced in the seventies, brought for a while an acceptable solution to this problem. Current trends show that we are clearly approaching the limits of this concept particularly in regard of the overload of pilot's short term memory. Anthropometric aspects have also to be considered and will definitely be aggravated with the arrival of female crew in the cockpit. Introduction of Alternative Controls constitutes a way to alleviate these problems. They also offer new possibilities, as for example the capacity to fire missiles with large off-boresight angle given by Helmet Mounted Sights. Last but not least, alternative controls pave the way towards the virtual cockpit, which may potentially result in substantial cost and space savings, as the need for many bulky and expensive control panels could be removed. This raises another question about the degree of maturity of the various Alternative Control Technologies.

The review of the different alternative technologies, which constitutes the core of this report, indicates quite clearly that there is large variation in terms of technical maturity of the concepts and devices used by the different alternatives. Voice control currently offers the best alternative way in regard of system management needs. Though current systems are quite satisfying for military operations, a large technical growth potential still exists for speech recognizers. This technology is also most susceptible to spread widely and grow in the public domain, which will provide additional thrust for the development of aerospace applications. The technology of head trackers used in HMDs seems now pretty mature and quite stabilized. Further enhancements will probably be necessary to exploit fully the possibilities of HMDs in virtual cockpits. Visually coupled systems (implicit control) are currently the main field of application of the head tracker technology. It has to be noted that for sighting applications (explicit control), head pointing is far from optimal from a human factors point of view. Detection of the line of gaze would allow use of the natural eye-head coordination mechanism which is currently disrupted in Helmet Mounted Sights. The eye-tracker technology required to build gaze-tracking devices can be considered as relatively mature in the R&D domain. Efforts should be made in regard of robustness, automation and integration in the cockpit before this technology becomes really usable in the aerospace environment. Potential applications of gaze tracking are numerous, from sighting to implicit monitoring of the pilot's state. Conversely, technologies using gesture and biopotentials are still clearly in the research domain. Gesture is a very basic and natural communication channel and is potentially very powerful. Unfortunately, the sensing devices are very often intrusive and may interfer with operators' activity. Progress in the miniaturisation of energy sources and electronic devices can help solve these problems. Substantial research remains to be done in the area of how to use and integrate such gesture-based systems. Biopotentials (EMG, EEG) are definitely a very fascinating and promising research domain. The biopotential technology could potentially lead to true « state/intent-based » systems, not requiring any kind of mechanical movement. Both signal processing and acquisition techniques remain to be considerably improved before reaching this goal and, with the exception of prosthetic device control, little work has been done so far outside the laboratory. Besides their advantages, limitations and challenges, either short term and long term Alternative Control Technologies share a key issue: System integration.

System integration issues can be split between two axes: human factors and system engineering considerations. Without sufficient attention to these domains, there is little chance that a successful integration could be achieved and potential benefits of Alternative Controls fully delivered.. Despite the knowledge individually accumulated on each technology, it has to be considered that the determination of the « good practice » to integrate these technologies is still in its infancy. More than conventional controls, Alternative Controls Technologies require a « human centred design » to fulfill the ultimate goal of designing a true « joint cognitive system ». Along with physiological and psychological knowledge, cognitive ergonomics can help to address the integration problems. The variability which characterizes the human being and the impact of imprecise and uncertain data relating to the field of alternative controls deserve an approach making good use of methods and tools developed by human factors scientists. Evaluation and performance measures should also be carefully conducted following appropriate human factor guidance. It is quite likely that current design guidelines available to system engineers will not be sufficient to cover all the integration issues. Subsequent to the introduction of novel controls, the physical arrangement of the cockpit and the flow of information between the cockpit and the remainder of the aircraft systems would have to be quite substantially reconsidered. This is expected to require innovative approaches and ingenuity from system design engineers.

The purpose of this report was to synthetically gather information on Alternative Control Technologies and to explore the issues and difficulties on the way to a practical integration of novel, non-conventional controls in aircraft cockpits. This work is based on the knowledge and diversified expertise of the members of the Working Group, completed by a thorough analysis of available literature. By no means can this work pretend to be an exhaustive review of all aspects pertaining to the vast domain of designing, building and using Alternative Control Technologies. As a matter of fact, this domain includes many application fields, such as virtual reality, computer interface, robotics,...., far beyond the scope of our group. Though centred on the aerospace environment, the information collected during the two years of activity of the Working Group should be of some use for other defense applications. It is hoped that the data and guidelines presented in this report will be helpful to scientist and engineers working in this area.

LIST OF ABBREVIATIONS

AC	Alternating Current
AFI	Acoustic Feature Identification
AGC	Automatic Gain Control
AMRL	(US Air Force) Aeromedical Research Laboratory.
APD	Acoustic Phonetic Decoding
ASR	Automatic Speech Recognition
ATC	Air Traffic Control
ATR	Air Transport Racking
AWACS	Airborne Warning And Control System
CofG	Centre of Gravity
CCD	Charge-Coupled Device
CEP	Circular Error Probable
COTS	Commercial Off-The-Shelf
CREAM	Cognitive Reliability and Error Analysis Method
dB	Decibels
dBA	Decibels (A-Weighted)
DC	Direct Current
DERA	Defence Evaluation and Research Agency
D/NAW	Day/Night All Weather
DOF	Degree-of-Freedom
DSP	Digital Signal Processing
DTW	Dynamic Time Warping
DVI	Direct Voice Input
EEG	Electroencephalographic or Electroencephalogram
EMG	Electromyographic or Electromyogram
ERD	Event-Related Desynchronisation
ERP	Event-Related Potential
ERS	Event-Related Synchronisation
FAC	Forward Air Controller
FEBA	Forward Edge of Battle Area
FFT	Fast Fourier Transform
FLIR	Forward-Looking Infra-Red
FOV	Field Of View
FOR	Field of Regard
GEC	General Electric Company (UK)
HF	High Frequency (communications)
	Human Factors (psychology)
HMD	Helmet Mounted Display
HMM	Hidden Markov Model
HOCAS	Hands On Collective And Stick
HOTAS	Hands On Throttle And Stick
HUD	Head-Up Display
I/O	Input/Output
IMELDA	Integrated Mel-scaled Linear Discriminant Analysis
LVQ	Learning Vector Quantization
MCDS	Multifunction Cathode Ray Tube Display System
MFCC	Mel Frequency Cepstral Coefficient
MFD	Multi-Function Display
MLP	Multi-Layer Perceptron
NASA	National Aeronautics and Space Administration
NBC	Nuclear, Biological and Chemical
NC	Noise Cancellation
NN	Neural Networks
PEC	Personal Equipment Connector
PLP	Perceptual Linear Prediction
PTT	Push-To-Talk
PWB	Perfect Word Boundary Detection
RAE	Royal Aircraft Establishment
RASTA	Relative Spectra
RMS	Root Mean Square
S/N	Signal-To-Noise
SD	Signal Detection
SNR	Signal-To-Noise Ratio
SPL	Sound Pressure Level
SRR	Sentence Recognition Rate
TDNN	Time-Delay Neural Network
TIALD	Thermal Imaging And Laser Designation
TIMIT	Texas Instruments/Massachusetts Institute of Technology
THERP	Technique for Human Error Rate Prediction
TMJ	Tempero-Mandibular Joint
UAV	Unmanned Air Vehicle
UHF	Ultra High Frequency
UK	United Kingdom
VEP	Visual-Evoked Potential
VHF	Very High Frequency
VQ	Vector Quantization
WPAFB	Wright Patterson Air Force Base
WRR	Word Recognition Rate

GLOSSARY

(Note that the connotations of a word may depend upon the context, and vary between chapters.)

CHAPTER 1

field of view -- The instantaneous solid angle over which a sensor or display is effective.

field of regard -- The total solid angle over which a steerable sensor or helmet-mounted display may operate.

SECTION 2.1

coarticulation -- The continuous movement of the speech articulators from the positions needed to produce one phone to those needed to produce the next. This means that there is no clear acoustic boundary between phones, and features (e.g. nasalisation) may spread into an adjacent phone, of which they would not otherwise be a part.

phoneme -- The smallest unit of speech sound capable of discriminating between words.

phone -- A unit of speech sound defined in terms of production. A phoneme may be produced in several ways, depending on context

diphone -- A segment of speech from the centre of one phone to the centre of the next (used mainly in speech synthesis).

perplexity -- A measure of the complexity of a grammar, usually calculated as the average of the number of words which may occur following each node in the syntax.

Mel scale -- A frequency scale based on perception of pitch, as opposed to physical frequency.

speaker dependent -- describes a speech recogniser which is trained separately for each user.

speaker independent -- describes a speech recogniser trained from a general population, that can be used by anyone without further training.

SECTION 2.2

AC coupled -- process that is correlated (coupled) with a periodic function. Knowledge of such correlation can be used to help distinguish the process from other signals that are not correlated to the same periodic function.

AC technique (for magnetic head tracking) -- magnetic tracking technique in which transmitter antennae are excited with sinusoidal currents, and induced currents of the same frequency are detected in a receiver (sensor).

compensatory tracking -- attempt to use a control input to stabilise a target that is being perturbed; in other words, to compensate for the perturbation.

DC technique (for magnetic head tracking) -- magnetic tracking technique in which transmitter antennae are excited with square current pulses (current values that remain constant during the duration of the pulse), and resulting induced currents are detected in a receiver (sensor).

explicit control -- purposeful control activity, for instance when using a helmet-mounted sighting system to aim at a target.

implicit control -- incidental control activity, for instance when using natural head motion which is sensed by the head tracker to slew the gimballed sensor in a visually-coupled system.

inertial package -- an enclosed unit containing several inertial sensors (such as rate gyros and accelerometers).

phase coherence (technique for acoustic tracking) -- a technique for detecting motion of a sound emitter (speaker) by comparing the phase of sound waves from the emitter to the phase of sound waves from a reference source.

root mean square -- a value generated by squaring the values of a process, finding the mean of the squared values, and taking the square root of this mean. It is a value sometimes used to describe the characteristic, or effective, magnitude of a varying process. For example, the root mean squared value of an alternating current is the direct current value that would cause the same amount of heat to be dissipated in a resistor.

triangulation -- use of geometric (or trigonometric) principles to find the position of the third point of a triangle when the positions of two points are known, and when the directions of the lines connecting those two points to the third point are also known.

SECTION 2.3

angular subtense -- The angle between (subtended by) two lines emanating from a common point.

cornea -- the clear (transparent), structure covering the iris and pupil of the eye.

electro oculogram -- recordings of the direction of the corneo-retinal potential (electric dipole between the cornea and retina of the eye) using signals from electrodes placed around the orbit of the eye.

eye tracker – device that measures (tracks) the orientation of the eye ball with respect to the measuring device.

fixation – period during which gaze is relatively stationary (fixed) with respect to some target image and visual information can be acquired and processed.

fovea (or fovea centralis) – a small roughly disk shaped area on the human eye retina (forming a slight depression within the yellowish spot called the macula), located about 2mm to the temporal side of the central optic axis of the eye, and having the most densely packed receptors, and therefore offering the highest visual acuity on the retinal surface.

gaze tracker -- device that determines a person's point of gaze or line of gaze with respect to the environment. This may sometimes require, for example, both an eye tracker to measure the orientation of the eye ball with respect to head

mounted optics, and a head tracker to measure the position and orientation of the head with respect to objects in the environment.

hot mirror -- material that is highly reflective to light in the infra red spectrum while being very transmissive to light in the visible spectrum.

limbus -- the ring shaped boundary on the eye ball where the iris, sclera, and cornea converge.

line of sight -- pointing direction of the visual axis of the the eye. Line of sight can be thought of as the infinite extension of the line beginning at the fovea and passing through the centre of the pupil.

macula -- a yellowish spot on the eye retina, located just to the temporal side of the central optic axis of the human eye, containing a slight central depression called the fovea (or fovea centralis).

mydriatic -- a drug that causes the eye pupil to dilate.

otoliths -- the sensors in the human inner ear, forming part of the structure called the vestibular system, that detect specific force (sum of gravitational and linear acceleration forces). The otoliths are formed by gelatenous masses called otoconea supported by sensory hair cells and supporting cells, and function in a fashion that is analogous to a mechanical spring mass system.

point of gaze -- the physical point whose image is intersected by the visual axis of the eye (the azis passing through the center of the pupil and intersecting the fovea). The point of gaze is imaged on the eye fovea (the high acuity area on the retina).

point of regard -- the part of the visual field that a person is attending to or "looking at". This is usually, but not always, the same as the point of gaze.

saccade -- rapid eye movement between fixation points during which visual gain is reduced and relatively little visual information is acquired. Peak eye ball rotation velocities often exceed 600°/sec during saccades.

sclera -- the opaque white structure that forms most of the outer surface of the eye ball. The sclera does not cover the iris and pupil, which are covered by the transparent cornea.

SECTION 2.4

epistemic - that pertains to perception.

deictic - intended to show or designate a specific object.

tracker - device allowing to measure the position and orientation of an object in real time, e.g. by mechanical means or using magnetic fields, ultrasonic or infrared beams.

pattern matching/recognition - comparing data to a set of patterns, in order either to find one present in the studied data (exact matching), or to decide, usually by computation of a similarity function, to which pattern it is most similar.

SECTION 2.5

Biofeedback - a method whereby normally unobservable physiological parameters, such as blood pressure, body temperature or brain electrical activity are displayed, in real time, to a user. Using biofeedback and structured training techniques, users can learn to regulate the physiological processes underlying these signals.

Brain rhythms - patterns in the electroencephalogram characterized by high power in a specific frequency band. Common examples are the alpha rhythm (8-12 Hz) and theta activity (4-7 Hz).

Common Mode Interference - an external electric field capacitively coupled to the body and appearing on both electrode inputs to a differential amplifier. By using differential amplification, the unwanted interference can be eliminated. Interference from the electrical mains is an example of common mode interference.

Delay - pure dead time in a system. Delay does not change the shape of an input waveform but simply shifts it in time. Transport delay is the more specific term.

Evoked Response - a measurable, characteristic change in brain electrical activity produced by some external stimulus, such as light, sound or touch.

Lag - a more complex form of delay produced by filter elements in a system. Because the effect of a filter depends on the frequency content of the input signal, it can distort the shape of an input waveform and shift it in time.

M-sequence - a white pseudo-random sequence of binary events, such as zeros and ones.

CHAPTER 3

explicit function - An operation performed by a system with sufficient participation by the user to maintain his knowledge of what is happening.

implicit function - An operation performed by a system without the involvement of the user.

cognitive task analysis - A delineation of the decisions made by individual in order to perform a task. This usually includes a description of the sources of information, and perhaps the means of implementing the choices made.

cognitive ergonomics - The study of work which requires thought or, more usually, work involving decisions.

servo paradigm - The mode of control in which the human operator interacts directly, and usually manually, with an effector system using a control device. The effector system responds consistently to the operator input.

structural coupling paradigm - A mode of control in which the human operator interacts indirectly with an effector system through a control system which monitors the overall user behaviour. In general the control system must sense the user's outputs, interpret his intentions with due regard for the instantaneous circumstances, and issue the direct control commands to the effector. The structural coupling paradigm is an alternative to the servo paradigm.

joint cognitive system - A term used to describe the combination of an operator and a machine which is more like a team than a master and slave. In general the machine should have access to reliable and relevant information and be able to make complex decisions.

hierarchical task analysis - The decomposition of a job performed by an operator into sequences of defined activities occupying defined durations, then the decomposition of these activities into more detailed actions, performed to a "granularity" (time interval resolution) which is sufficient to show relevant phenomena.

taxonomy - A mutually exclusive set of classes or categories into which items can be usefully separated.

Air Transport Racking -- A specification for the dimensions and the mechanical, electrical and environmental attributes of avionic equipment.

Personal Equipment Connector -- A term used by a manufacturer of escape systems to denote the quickly-detachable device used to couple the electrical cables, the breathing air and the g-suit inflation air to man-mounted equipment from the aircraft systems through a linking plate fixed to the ejection seat.

command interpreter - A means of adjudicating between the outputs of separate control systems, either conventional or novel, in order to send an unambiguous command to the relevant aircraft system.

transaction - A set of operator/system actions which must be completed to be effective, i.e. the entry of a complete data string needed to specify the latitude and longitude of a waypoint.

inceptor - A term coined by control engineers to signify any device which the operator uses to generate an input to a controlled system.

deceptor - A term coined by control engineers to signify a device which is too unreliable for an operator to use to generate an input to a controlled system.

Wizard-of-Oz technique -- An experimental paradigm in which the subject is led to believe that he or she is communicating with a machine, but in fact a hidden human operator is producing the responses.

APPENDIX A
WAVELETS

The Continuous Wavelet Transform of a signal $s(t)$ is defined by:

$$S_g(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) g\left(\frac{t-b}{a}\right) dt$$

where b is a duration parameter, a is a scale factor and $g(t)$ is a function called an analysing wavelet and has the following properties:

1. The energy of $g(t)$ is finite : $\int_{-\infty}^{\infty} |g(t)|^2 dt < +\infty$

2. The average of $g(t)$ is zero: $\int_{-\infty}^{\infty} g(t) dt = 0$

The main objective of Wavelet Analysis is to point out non-stationary short duration phenomena. So, the Wavelet Transform must not extract particular features of the signal *s(t)=1* . So, we must have :

$$S_g(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} g\left(\frac{t-b}{a}\right) dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} g(t) dt = 0$$

with *s(t)=1*.

3. Admissibility condition: $\int_{-\infty}^{\infty} |\hat{g}(\omega)|^2 \frac{d\omega}{\omega} < +\infty$

where $\hat{g}$ is the Fourier Transform of $g(t)$

This condition shows the reconstruction formula which will be described.

The $S_g(b,a)$ values are called the wavelet coefficients of $s(t)$ associated with the analysing wavelet $g(t)$. Then, it can be shown that the scale factor controls the time-frequency resolution.

As said before, the admissibility condition allows to show the reconstruction formula:

$$s(t) = \frac{1}{\sqrt{c_g}} \int_{a,b} \frac{1}{\sqrt{a}} S_g(b,a) g\left(\frac{t-b}{a}\right) \frac{1}{a^2} da\, db$$

$$\text{where } c_g = \int_{-\infty}^{\infty} |\hat{g}(\omega)|^2 \frac{d\omega}{\omega}$$

The wavelet coefficients can be computed for discrete values of a and b through the A TROUS Algorithm [A-1] which can be considered a fast algorithm. In [A-2], it is shown that the values of a and b can be taken equal to:

$$a = a_0^m \text{ and } b = k b_0 a$$

without any loss of the signal to analyse.

These very important results show that the Wavelet Transform can be computed through discrete algorithms.

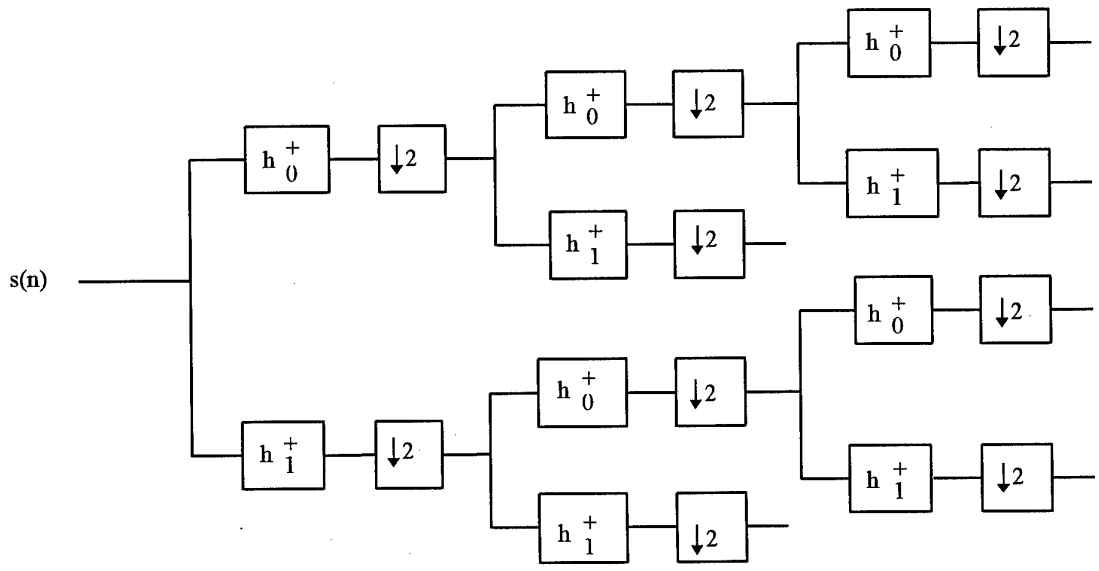


Figure A- 1 Example of orthonormal wavelet packet decomposition.

↓2 represents the decimation operator which eliminates alternate samples

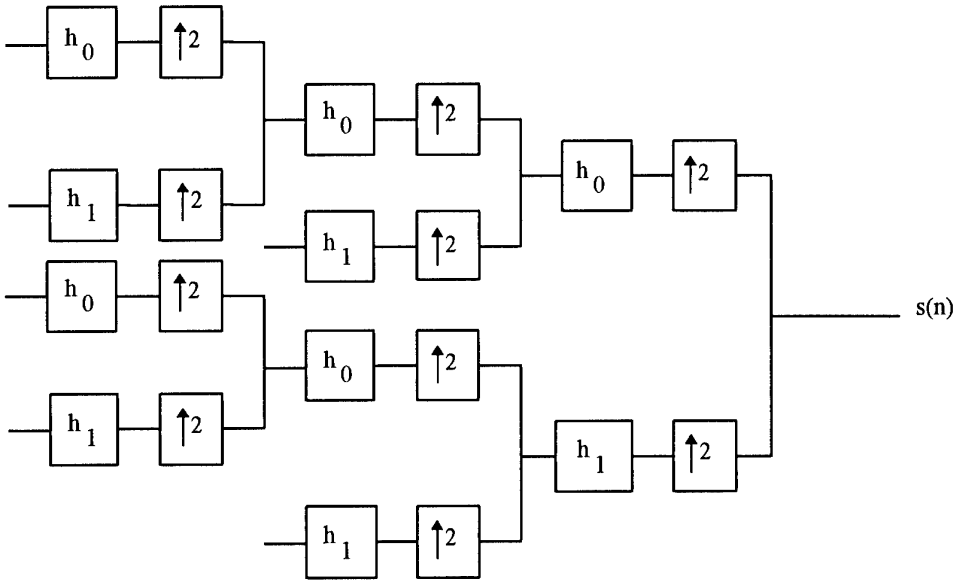


Figure A- 2 Example of orthonormal wavelet packet reconstruction.

↑2 represents the discrete dilatation operator which inserts a null sample between two samples

It is then very interesting to have a look at the function

$$g_{a,b}(t) = \frac{1}{\sqrt{a}} g\left(\frac{t-b}{a}\right)$$

when using $a = a_0^m$ and $b = kb_0 a$. It easily comes: $g_{a,b}(t) = a_0^{-m/2} g\left(a_0^{-m} t - kb_0\right)$. The values $a_0 = 2$ and $b_0 = 1$ are often chosen and we write: $g_{m,k}(t) = 2^{-m/2} g\left(2^{-m} t - k\right)$. Then, taking $g_{m,k}(t)$ as Orthonormal Bases of Finite Energy Signals, leads to Discrete Orthonormal Wavelet Analysis [A-3] and Discrete Wavelet Packets Analysis [A-2, A-4] which can be considered as orthonormal filter banks.

Discrete Orthonormal Wavelet Packets are very efficient from a practical point of view, because the analysis of a signal is a recursive one. So, such a decomposition will be implemented through a tree structure as shown in Figure A-1, where:

$$h_0^+(n) = h_0(-n)$$

h_0 and h_1 are two mirror quadrature filters, which means that their respective Fourier Transform, $m_0(\omega)$ and $m_1(\omega)$ verify the following equations:

- $|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 2$
- $|m_1(\omega)|^2 + |m_1(\omega + \pi)|^2 = 2$
- $m_0(\omega)\overline{m_1(\omega)} + m_0(\omega + \pi)\overline{m_1(\omega + \pi)} = 0$

One of the main advantages of this analysis is to provide a reconstruction algorithm whose structure is quite the same as the decomposition one. For example, the reconstruction of the sampled signal analysed through the previous decomposition, will be carried out as shown in Figure A-2.

The last technique we must mention is the Malvar Wavelet Transform which can be considered as a windowed Fourier Transform over lapped intervals.

All these wavelet techniques are supported by a tremendous amount of mathematical tools which allow efficient theoretical and practical developments, while generalizing older methods.

REFERENCES

A-1 R. Kronland-Martinet. "Analyse, synthèse et modification de signaux sonores: application de la transformée en ondelettes", Thèse de la faculté des sciences de l'Université d'Aix-Marseille II, 1989.

A-2 I. Daubechies, "Ten Lectures on Wavelets", Society for Industrial and Applied Mathematics, 1992, Philadelphia

A-3 S. Mallat, "A theory for Multi-Resolution Signal Decomposition: the Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, no. 7, 1989.

A-4 V. Wickerhauser, "Adapted Wavelet Analysis from Theory to Software", A. K. Peters, 1994, Massachusetts.

APPENDIX B
DYNAMIC TIME WARPING

As explained in Section 2.1.2.3, it is necessary to find the optimum correspondence between the vectors of the incoming speech and those of each model in the active vocabulary. In its simplest form, as described here, the DTW algorithm assumes that the start and end points of the input word are known. The input word and the model will in general consist of different numbers of frames. A linear correspondence in time between the frames of the word and those of the model cannot be assumed.

After being processed by the front end, the unknown word will be represented by a sequence of m p-dimensional observation vectors:

$$X = \{O_X(1), O_X(2) \ldots\ldots O_X(m)\}$$

where

$$O_X(i) = \{x_1, x_2, \ldots\ldots x_p\}$$

The k^{th} model will be represented by a similar sequence of length n:

$$Y_k = \{O_{Yk}(1), O_{Yk}(2), \ldots\ldots O_{Yk}(n)\}$$

where

$$O_{Yk}(j) = \{y_{k1}, y_{k2}, \ldots\ldots y_{kp}\}$$

The first step in the DTW algorithm is to calculate the local distance matrix between each vector of the input word and each vector of the model. The distance metric used will be dependent on the signal representation; the squared Euclidean distance is often used with a filter bank output:

$$D_k(i,j) = \sum_{q=1}^{p} \left[x_i(q) - y_{kj}(q)\right]^2$$

where $D_k(i,j)$ is the distance between the i^{th} vector of the input word and the j^{th} vector of the k^{th} model. Figure B-1 represents the local distance matrix for comparing a five state input word (m = 5) with a seven state model (n = 7). The first state of each word is at the bottom left corner; the sequence of subsequent states is from left to right for the input word and from bottom to top for the model. The problem then becomes one of finding the optimum path from the bottom left corner to the top right corner of this matrix. The optimum path is defined as that which accumulates the lowest total distance score.

$D_k(1,n)$	..	..	..	$D_k(m,n)$
..	..	..	..	..
..	..	..	$D_k(i,j)$	..
..	..	..	..	..
..	..	..	..	..
$D_k(1,2)$	$D_k(2,2)$	..	..	..
$D_k(1,1)$	$D_k(2,1)$	..	..	$D_k(m,1)$

Figure B-1 Local Distance Matrix

The next step is to calculate the cumulative distance matrix. This is generated from the local distance matrix by a recurrence relationship:

$$g_k(i,j) = \min \begin{bmatrix} g_k(i-1,j) + D_k(i,j), \\ g_k(i-1,j-1) + WD_k(i,j), \\ g_k(i,j-1) + D_k(i,j) \end{bmatrix}$$

The cumulative distance at any point in the matrix is the sum of the local distance and the minimum of the cumulative distances of the neighbouring elements in the matrix. In this case, transitions to any cell are allowed from the left, below or low-left diagonal, as shown in Figure B-2(a). The weighting factor W may be applied to the diagonal transition to compensate for the smaller number of local distances accumulated on this path. The optimum value for W is not known; values of 1 or 2 are generally used.

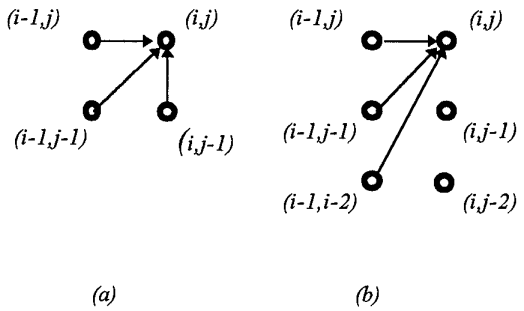


Figure B-2 Examples of Allowed Transitions

This is an example of the Viterbi algorithm [B-1] which has wide application in pattern matching and signal decoding. Many variants are possible within the same basic framework, for example, the use of asymmetric transitions as shown in Figure B-2(b).

When the cumulative distance matrix is complete, the value in the top right hand corner, $g_k(m,n)$ is the total cumulative distance between the input word and the model, along the

optimum time registration path. This value is normalised before use, to compensate for different model lengths:

$$G_k(m,n) = \frac{g_k(m,n)}{(m+n)}$$

The model which gives the lowest value of $G_k(m,n)$, i.e. is nearest to the input word in this p-dimensional space, is taken to represent the word which was spoken.

In a real application, the observation vectors will consist of about twenty elements, and the input word and model may consist of about 30 frames each. Calculating the local distance matrix will therefore require on the order of 30 x 30 x 20 = 18,000 multiply operations *for every model.* This computational load places one of the main constraints on active vocabulary size. This can be alleviated to some extent by recognising that there are limits on the extent to which the time registration path is likely to deviate from the diagonal, so some cells in the top left and bottom right corners of the distance matrices need not be calculated, Figure B-3.

x	x	x	x	x	x				o
x	x	x	x	x					o
x	x	x	x					o	
x	x	x				o	o		
x	x				o				x
x					o			x	x
					o		x	x	x
			o	o		x	x	x	x
		o			x	x	x	x	x
o	o			x	x	x	x	x	x

Figure B-3 Unused Areas of Distance Matrix
o = optimum time registration path
x = Area unlikely to be used

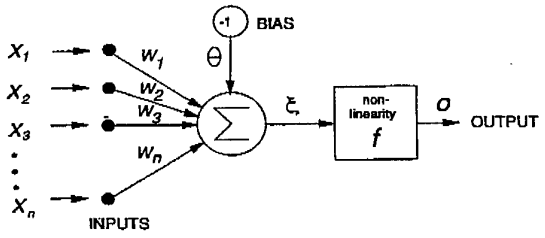
The basic DTW algorithm as described above is best suited to work with simple models, such as single utterance templates.

REFERENCES

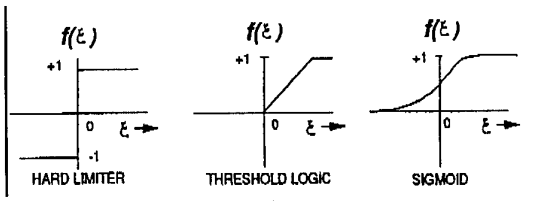
B-1 Viterbi, A. J. "Error bounds for convolution codes and an asymptotically optimal decoding algorithm." IEEE Trans. on Information Theory, IT-13 1967 pp 260-269

APPENDIX C
ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) are a family of computing systems inspired by the learning capability and structure of biological neural networks. ANNs have been used for many purposes, including signal filtering, data compression, time series prediction, optimization problems, and as content addressable memories, (see review in [C-1]). Their structure and training as a method of pattern classification are outlined briefly in this section.



(a)



(b)

Figure C- 1(a) An artificial neuron. (b) Three typical neuron non-linearities, (adapted from [C-1]).

The basic element of an ANN is shown in Figure C-1a. For this single artificial neuron, the activation ξ is the sum of all inputs x_i multiplied by connection weights w_i , minus a threshold θ. The activation is passed through a non-linearity $f(\xi)$ to produce the neuron output o:

$$o = f(\xi) = f\left(\sum_{i=1}^{n} w_i x_i - \theta\right)$$

The non-linearity can have many forms including those shown in Figure C-1b. This artificial neuron can divide the input feature space into two regions separated by a boundary which is dependent on the form of the non-linearity, (e.g. if a hard limiter non-linearity is used, a simple hyperplane is formed). The feature space can be partitioned further by adding more neurons to the network.

For pattern classification, the network of neurons is usually trained by adjusting the connection weights such that an input pattern from a given class causes one neuron output to be high, while the others remain low. Several algorithms have been developed to determine a set of weights and a neuron threshold which can correctly classify a set of training patterns. This can be done in a supervised mode, in which the desired output is available to the network during learning, or in an unsupervised mode, in which clusters are formed from the input patterns. Supervised learning can be done through an iterative procedure based on the delta rule. The delta rule states that the change in the weight Δw_i connecting neuron j with input x_i is proportional to the product of the difference between the desired neuron output d and the actual neuron output o and the value of the input feature:

$$\Delta w_i = w_i(new) - w_i(old) = \eta\,(d - o)x_i$$

where η is the learning rate of the procedure. Care must be taken to select a learning rate which is high enough to achieve a satisfactory rate of weight adaptation but does not produce weight instability.

In a typical application, the network determines the output in response to an input training pattern. This actual output is compared to the desired output for that input pattern, to determine the output error. If the error exceeds a specified value, the weights are updated by adding a fraction of the misclassified pattern to the weight vector. Typically many representative patterns from each classs form the training data set. It is necessary to cycle through the training data many times until all patterns are classified correctly or some other stopping criterion is met.

The basic delta rule works well for single layered networks (only input neurons and output neurons). Single layer networks, however, are capable of solving only linear classification. This was a major limitation on the use of neural networks for practical pattern classification problems. A generalized delta rule known as backpropagation has been developed which is capable of training multi-layered networks, [C-2, C-3]. The backpropagation algorithm provides a means of updating the network weights in the middle layers (know as the hidden layers) of a network by using the output errors from a higher layer. It has been shown that a network with one or more hidden layers can realize arbitrary continuous mappings from input to output.

In general, an ANN is made up of sets of neurons arranged in layers as shown in Figure C-2. In this figure and in the subsequent derivations the superscript in parenthesis denotes the network layer being considered. The neuron outputs from one layer are used as inputs to the next layer after being attenuated or amplified by a set of weighting factors. The pattern features are the inputs to layer I. The output of this layer may be the original feature values or, if necessary, the inputs may be scaled to ensure that each is within some reasonable range. No activation function is associated with the input nodes.

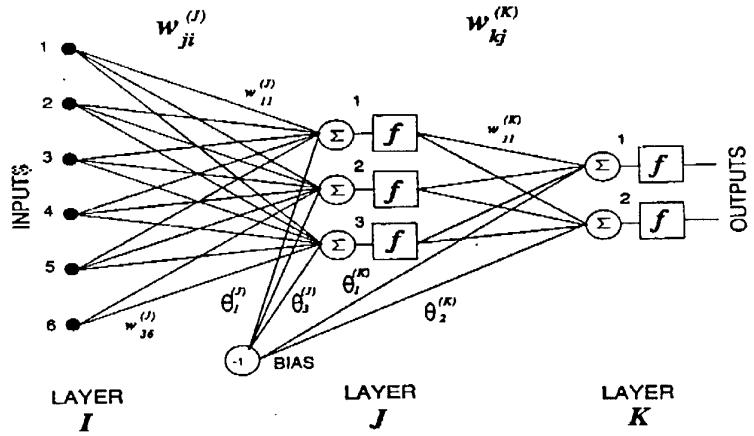


Figure C-2 A fully connected artificial neural network with one hidden layer.

For a fully connected network in which all outputs from layer I are connected to all inputs of layer J , the net input to neuron j in layer J is:

$$\xi_j^{(J)} = \sum_{i=1}^{n} w_{ji}^{(J)} o_i^{(I)} - \theta_j^{(J)}$$

where $\theta_j^{(J)}$ is the threshold for neuron j, w_{ji} is the connection weight between the output of neuron i in layer I and input of neuron j in layer J, and the summation is taken over all n outputs from layer I. The input layer I has no activation function therefor the n outputs of this layer $o_i^{(I)}$ are the input features.

Assuming a sigmoidal activation function for the neurons of layer J, the output of neuron j in layer J is computed as:

$$o_j^{(J)} = \frac{1}{1+\exp(-\xi_j^{(J)})} \qquad (1)$$

Likewise the corresponding input for neuron k in layer K is:

$$\xi_k^{(k)} = \sum_{j=1}^{n} w_{kj}^{(K)} o_j^{(J)} - \theta_k^{(K)}$$

where the summation is taken over all m outputs of layer J.

Again assuming a sigmoidal activation function for the neurons in layer K, the output of neuron k is:

$$o_k^{(K)} = \frac{1}{1+\exp(-\xi_k^{(K)})} \qquad (2)$$

Backpropagation is a supervised training algorithm which minimizes the mean squared error between the actual output and the desired output for a set of input/output training patterns. For each input pattern this error is given by:

$$E = \frac{1}{2}\sum_{k=1}^{q}(d_k^{(K)} - o_k^{(K)})^2 \qquad (3)$$

where q is the total number of neurons in layer K.

The error is a function of all network weights and thresholds. A threshold can be considered as a fixed input of -1 connected to the neuron through a weight which is adapted in the same manner as other weights. Convergence is achieved by making incremental changes to the network weights proportional to their accountability for the error. For each network weight $w_{yx}^{(Y)}$ connecting the output of neuron x in layer $Y-1$ to the input of neuron y in layer Y, the change in weight is defined as:

$$\Delta w_{yx}^{(Y)} = -\eta \frac{\delta E}{\delta w_{yx}^{(Y)}} o_x^{(Y-1)}$$

The partial derivative can be expanded using the chain rule. Details of the exact derivation are given elsewhere [C-4]. This results in weight changes which are a function of only local outputs and of weight changes from higher layers. Specifically, for a sigmoidal activation function, the weight change for the connection between neuron k in layer K and neuron j in layer J is given by:

$$\Delta w_{kj}^{(K)} = -\eta \delta_k^{(K)} o_j^{(J)} \qquad (4)$$

where

$$\delta_k^{(K)} = (d_k^{(K)} - o_k^{(K)}) o_k^{(K)} (1 - o_k^{(K)})$$

The weight change for the connection between neuron j in the hidden-layer and input node i is given by:

$$\Delta w_{ji}^{(J)} = -\eta \delta_j^{(J)} o_i^{(I)} \quad (5)$$

where

$$\delta_j^{(J)} = o_j^{(J)}(1 - o_j^{(J)}) \sum_{k=1}^{q} \delta_k^{(K)} w_{kj}^{(K)}$$

and where the summation is taken over all q neurons in layer K.

The network is trained by first initializing all weights and thresholds to small random values. An input/output pair is presented to the network. This produces an output for each neuron as calculated by Equations 1 and 2 in the feedforward direction. The error for this training pattern is calculated at the output neurons and the weights for the neurons in the output layer are updated using Equation 4. These updates are then back propagated to the hidden layer so that the weight updates for the hidden layer neurons can be calculated using Equation 5. The next input/output pair is presented and the actual output is calculated using the updated weights. The training pairs are cycled until the weights stabilize or until the output error (calculated using Equation 3) averaged across all training pairs is below a desired threshold.

Given enough training data and a well chosen feature set, the hidden layer forms appropriate general purpose feature detectors or filters which allow the network to classify novel inputs. The hidden layer also gives a network the potential for an arbitrary mapping of the input features through a nonlinear transformation onto the domain of the hidden layer output. This domain is partitioned by the output neurons to determine class assignment. The optimum number of hidden units has not been established and is problem dependent. Reducing the size of the hidden layer not only reduces the computational complexity of the neural network but can improve the network's ability to generalize from the training data. Too many hidden units will allow the network to memorize each of the patterns in the training set and will inhibit generalization. On the other hand, a network with too few hidden units will be unable to learn the necessary input/output mapping.

Backpropagation has been found to perform well in many pattern classification problems including speech synthesis and recognition, sonar target classification, handwritten character analysis, and automatic target recognition, (see reviews in [C-5, C-6]). Although the gradient descent nature of this algorithm means that learning can be slow, the network invariably converges to a good solution assuming good training data and adequate network size. Numerous modifications have been proposed to speed learning and to enhance the network performance. The primary limitations of this learning algorithm are the need for a large set of training patterns and the existence of many local minima on the error surface. These and other shortcomings have led to alternate network structures and learning strategies. Lippman [C-5], Hinton [C-6] and Hush and Horne [C-7], review many of these developments.

References

C-1 Lippman, R.P. "An introduction to computing with neural nets", IEEE ASSP Mag. April, 1987, pp 4-22.

C-2 Rumelhart, D.E. and McClelland, J.L. "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", (Vol. I), Cambridge, Ma., MIT Press, 1986, (ISBN 0-262-18120-7).

C-3 Werbos, P. "Beyond regression: New tools for prediction and analysis in the behavioural sciences", PhD dissertation, Harvard University, 1974.

C-4 Pao, Y.H. "Adaptive Pattern Recognition and Neural Networks", Addison-Wesley, Reading, Mass., 1988.

C-5 Lippman, R.P. "Pattern classification using neural networks", IEEE Communications Mag., November, 1989, pp. 47-64.

C-6 Hinton, G.E. "Connectionist learning procedures", Artificial Intelligence, Vol. 40, 1989, pp 185-234.

C-7 Hush, D.R. and Horne, B.G. "Progress in Supervised Neural Networks, IEEE Signal Processing Mag., January 1993, pp 8-39.

APPENDIX D
HIDDEN MARKOV MODELS

Template-based ASR methods, such as DTW, do not explicitly rely on the statistical characteristics of the speech signal. We now consider pure statistical approaches based on stochastic processes, especially Markov processes.

A Markov chain consists of a set of states, with transitions between the states. Each state corresponds to a symbol, and to each transition is associated a probability. Symbols are produced as the output of the Markov model by the probabilistic transitioning from one state to another. Such models can thus be used to study phenomena in which deterministic observed symbols are arranged in temporal series. However, this model is too restrictive to study complex problems like the ones associated with speech recognition. For that purpose, it is necessary to extend the model to be able to treat the case where the observations are probabilistic functions of the states. This yields a dual formulation: speech observations can be generated from states or transitions. The resulting Hidden Markov Model (HMM) will be briefly described in this section. More details can be found in [D-1 - D-4].

An HMM is similar to a Markov chain, except that the output symbols are probabilistic: in fact, all symbols are possible at each state, each with its own probability. Therefore, to each state is associated a probability distribution of all possible symbols. In other words, an HMM is composed of a non-observable "hidden" process (a Markov chain), and an observation process which links the acoustic vectors extracted from the signal to the states of the hidden process. In that sense, an HMM is a so-called doubly stochastic process.

Figure D-1 shows a five-state HMM representing a speech unit (phoneme, word, etc.) with the allowed transitions. This graph can be seen as a production model in which each transition corresponds to the emission of a speech frame or feature vector. To each state s_j corresponds a probability distribution $P(e_k|s_j)$ (probability of producing event e_k, when a transition from this state occurs), and to each arc a probability $a_{ij} = P(s_j|s_i)$ (probability of transition from state i to state j). Since there are strong temporal constraints in speech, left-to-right HMMs are generally used.

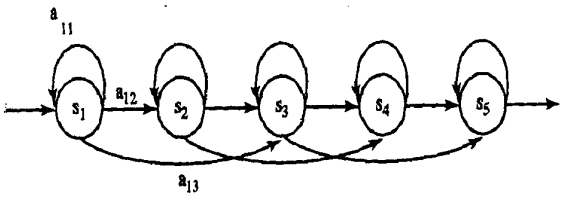


Figure D-1 Five-state left-to-right hidden Markov model

An HMM can model a specific speech unit such as a subword unit, a word, or a complete sentence. In large vocabulary recognition systems, HMMs usually represent subword units, such as phonemes, to limit the amount of training data and storage required for modeling words. Conversely, in small vocabulary systems, the tendency is to use HMMs to model words.

For any kind of HMM, three problems must be solved:

- *the evaluation problem:* given a model and a sequence of observations on the speech signal, what is the probability of the observation sequence, conditioned on the model. An efficient solution can be found with the forward pass of the forward-backward algorithm [D-1 and D-6];
- *the learning problem:* given an HMM and a sequence of observations, how to adjust the model parameters to maximize the probability of generating the observations (Maximum Likelihood (ML) criterion)? The observation sequence is called the training sequence. The training phase is crucial in the design of an HMM-based ASR system, since it makes it possible to optimally adapt model parameters to real-world phenomena. The learning problem can be solved using an iterative procedure such as the Baum-Welch algorithm (e.g. [D-5 - D-6]), a specific instance of the EM (Expectation-Maximization) algorithm.
- *the decoding problem:* given a model and a sequence of observations, what is the state sequence in the model that best explains the observations? The solution of this problem requires an optimality criterion to find the best possible solution. Typically, the Viterbi algorithm [D-7, D-8] is used. In the case of continuous speech recognition or subword unit systems, HMMs can be concatenated and the Viterbi algorithm finds the best model sequence corresponding to the observation data.

The remaining part of this subsection provides some mathematical details on how these problems are solved in the ASR context. To define an HMM model the following notations have been adopted:

$O = (O_1, O_2, \ldots, O_T)$ is a sequence of speech observations;

N is the number of states;

a_{ij} is the probability of transition from state i to state j;

$b_i(O_t)$ is the probability of state i emitting output vector O_t;

λ represents the set of parameters defining the HMM model (the transition probabilities A, the emission probabilities B, and the initial state distributions π).

The evaluation problem can be solved using the forward pass of the forward-backward algorithm. It is an efficient procedure to calculate $P(O|\lambda)$, the probability of the observation sequence O, given the model λ. Let the forward variable, $\alpha_t(i)$, be defined as

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_t, i_t = q_i|\lambda)$$

i.e. the joint probability of observing the first t speech vectors and being in state q_i at time t, given the model λ. $\alpha_t(i)$ can be calculated using the following recursion:

$$\alpha_{t+1}(j) = \left[\sum_{t-1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \quad \begin{matrix} 1 \le t \le T-1, \\ 1 \le j \le N, \end{matrix}$$

with the following initial conditions:

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \le i \le N.$$

Then, $P(O|\lambda)$ is given by

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)..$$

For a given time t, the computation of the forward variable, $\alpha_t(i)$, is performed for all states. Then it is iterated for t=1, 2.... T-1. This is the forward pass of the forward-backward algorithm. It is sufficient to calculate $P(O|\lambda)$. The combination of the forward and backward passes is used to solved the learning problem mentioned above. As with the computation of the forward variable, $\alpha_t(i)$, let consider the backward variable, $\beta_t(i)$, defined as

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T | i_t = q_i, \lambda),$$

i.e. the *conditional* probability of observing the speech vectors from *t+1* to the end and being in state q_i at time t, given the model λ. Similarly to the computation of the forward variable, $\beta_t(i)$ can be calculated using the following recursion:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad \begin{matrix} T-1 \le t \le 1, \\ 1 \le i \le N, \end{matrix}$$

with the following initial conditions:

$$\beta_T(i) = 1, 1 \le i \le N.$$

Note that at $\alpha_t(i)$ is the joint probability of arriving in state q_i at time t and observing the first t speech vectors, while $\beta_t(i)$ is the conditional probability of observing the last T-t vectors given that the state at time t is q_i. This asymmetry permits the calculation of the likelihood of state occupation by taking the product of the forward and backward variables.

To adjust the model parameters λ in order to maximize the probability of the observation sequence given the model, it is useful to define the following variable:

$$\gamma_t(i, j) = P(i_t = q_i, i_{t+1} = q_j | O, \lambda),$$

i.e. the probability of taking the transition from state i to state j at time t, given the observation sequence O and the model λ. $\lambda_t(i, j)$ can also be written as

$$\gamma_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)}$$

where $P(O|\lambda)$ is a normalization factor. Then, the expected number of transitions made from state q_i is $\sum_{t=1}^{T} \sum_{j=1}^{N} \gamma_t(i, j)$, and the expected number of transitions from state q_i to state is $\sum_{t=1}^{T} \gamma_t(i, j)$. Using the above formulas and the concept of counting occurrences, the probability of taking the transition from state i to state j can be re-estimated with the following equation:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T} \gamma_t(i, j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \gamma_t(i, j)}$$

Similarly, $b_j(k)$ can be re-estimated as the ratio between the frequency that symbol k is emitted in state j, and the frequency that any symbol is emitted in state j. This yields the following formula:

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T} \sum_{j=1}^{N} \gamma_t(i, j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \gamma_t(i, j)}$$

The two previous equations are known as the Baum-Welch re-estimation formulas. Every re-estimate is guaranteed to increase $P(O|\lambda)$, unless a critical point is already reached. In this case, the re-estimate will remain the same or even decrease if there is overtraining.

The decoding task can be solved by means of the Viterbi algorithm. It consists of matching an unidentified sequence of observation vectors, $O = (O_1, O_2, \ldots, O_T)$, against each of the models available. To find the single best state sequence for the first t observations, the best score, $\delta_t(i)$, along a single path, at time t, ending in state i, has to be defined:

$$\delta_t(i) = \max P[q_1, q_2, \ldots, q_{t-1}, q_t = i, O_1, O_2, \ldots, O_t | \lambda)$$

$\delta_t(i)$ can be computed using the following recursion:

$$\delta_t(j) = \begin{matrix} [\max \delta_{t-1}(i) a_{ij}] b_j(O_t), & 2 \le t \le T, \\ 1 \le i \le N & 1 \le j \le N, \end{matrix}$$

with the following initial conditions:

$$\delta_1(i) = \pi_i b_i(O_i), 1 \le i \le N.$$

The best state sequence for the observation vectors O can be obtained by keeping track of the argument that maximized $\delta_t(j)$ for each t and j.

The mathematical formulations presented above considered the case where the observation sequence belongs to a set of discrete symbols. However, the above solutions for the three HMM problems can be extended to cases where the observations are continuous multi-dimensional vectors. More details about the evaluation, learning, and decoding of HMM problems as well as implementation issues, such as statistics initialization and probability scaling, can be found in [D-1, D-6, D-9 - D-13].

References

D-1 Rabiner, L., and Juang, B.-W., "An introduction to hidden Markov models" , IEEE ASSP Magazine, 3(1), 1986, pp 4-16.

D-2 Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition" , Proc. IEEE, 77(2), 1989, pp 257-286.

D-3 Rabiner, L., and Juang, B.-W., "Hidden Markov models for speech recognition - strengths and limitations" , in P. Laface and R. De Mori, (Eds) "Speech Recognition and Understanding. Recent Advances, Trends and Applications", Springer-Verlag, 1992, pp 3-29.

D-4 Schwartz, R., and Kubala, F., "Hidden Markov models and speaker adaptation", in P. Laface and R. De Mori, (Eds) "Speech Recognition and Understanding. Recent Advances, Trends and Applications" , Springer-Verlag, 1992, pp 31-57.

D-5 Baum, L., "An inequality and associated maximization technique in statistical estimation for probabilistic functions for Markov processes", Inequalities, No. 3, 1972, pp 1-8.

D-6 Bahl, L., Jelinek, F. and Mercer, R., "A maximum likelihood approach to continuous speech recognition", IEEE Trans. On Pattern Analysis and Machine Inteligence, PAMI-5(2), 1983, pp 179-190.

D-7 Viterbi, A., "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm", IEEE Trans. On Information Theory, IT-13, 1967, pp 260-269.

D-8 Forney, G. D., "The Viterbi Algorithm", Proc. IEEE, Vol. 61, 1973, pp 268-278.

D-9 Baker, J., "Stochastic Modeling for automatic speech understanding", in R. Reddy (ed), "Speech Recognition", Academic Press, 1975, pp 521-542.

D-10 Jelinek, F., "Continuous speech recognition using statistical methods", Proc. IEEE, 64(4), 1976, pp 532-556.

D-11 Levinson, S., Rabiner, L., and Sondhi, M., "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition", Bell Sys. Tech. J., 62(4), 1983, pp 1035-1074.

D-12 Lee, K.-F., "Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System", Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, 1988.

D-13 Rabiner, L., and Juang, B.-W., "Fundamentals of Speech Recognition", Englewood Cliffs, NJ, Prentice-Hall, 1993. (ISBN 0 13 015157 2)

APPENDIX E
REFERENCE FRAME RELATIONSHIPS

E.1. GENERAL CONCEPT

A reference frame is an imaginary, 3 dimensional spatial framework that is rigid and is usually considered to be rigidly attached to some object or entity. Position information in a particular reference frame may be expressed as Cartesian coordinates, as spherical coordinates, or in terms of some other equivalent coordinate system.

Many of the human/machine interaction techniques under consideration require that information acquired with respect to one spatial reference frame be used for control or designation in a different frame. Often these frames are moving relative to one another, and in some cases their relative position or orientation is not well known.

For the purpose of the following discussion, reference frames will be defined in terms of right handed Cartesian coordinates. The location of any point in such a frame can be expressed as set of 3 coordinate values (x, y, and z). The point at which the three coordinate values are all zero is the "origin" of the frame.

At any instant every reference frame has a position and orientation with respect to every other reference frame. The position of frame *a* with respect to frame *b* is expressed as the position of the frame *b* origin, in the frame *a* coordinate system. The orientation (or attitude) of frame *a* with respect to frame *b* can be expressed in terms of three Euler angles, a cosine matrix, or a set of quaternions. (Mathematical definitions for Euler angles, cosine matrices and quaternions are presented in the section titled *Mathematical notations and conventions*.)

Positions or vectors specified in one coordinate frame can easily be transformed to equivalent coordinates with respect to a different frame if the relative position and orientation of the two frames are known. Trigonometric operators and matrix multiplication are usually needed; so if a coordinate transformation must be performed repetitively in real time, processing requirements must be considered.

E.2 REFERENCE FRAMES FOR HUMAN INTERACTION WITH AEROSPACE SYSTEMS

Reference frame definitions are arbitrary, and are generally established for best computational convenience. The following are examples of reference frame definitions that may sometimes be applicable to human interaction with aerospace systems. Some of these reference frames are also shown schematically in Figure E-1.

- A navigational reference frame is traditionally defined to be centered at the vehicle, with the *x* axis pointing north, the y axis pointing east and the *z* axis pointing straight down.
- An aircraft reference frame is traditionally defined to have the same center as the navigation frame but fixed to the aircraft such that the x axis is aligned with the aircraft longitudinal axis (points out of the nose), the y axis points out the right wing, and the z axis points down.
- The position of fixed objects within the cockpit (instrument panels, controls, etc.) may be known with respect to a cockpit reference frame, having some fixed, central position in the cockpit.
- Any panel in the cockpit may be considered to have a reference frame attached to it, and the location of individual instruments, displays, or controls on the panel may be specified with respect to this frame.
- Each instrument, display, or control may also have its own defined reference frame for specifying position of elements within the instrument.

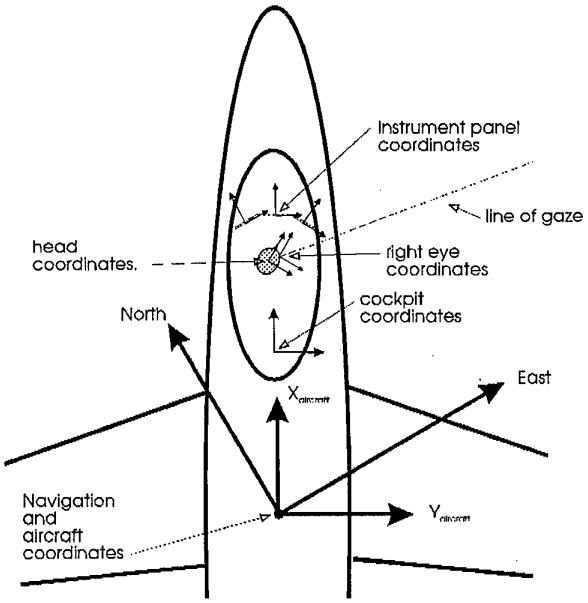


Figure E-1 Examples of different reference frames

- The position and orientation of the pilot's head is typically represented by a reference frame with its origin at the center of the head, and the x axis extending out from the front of the face.
- It may sometimes be convenient to describe eye line of gaze in terms of eye reference frames that are head fixed and have the same orientation as the head frame, but have origins at the center of each eye.
- A reference frame attached to the pilot's head gear might be used to specify the position and orientation of the head gear with respect to the pilot's head, and also to specify the positions of other head gear mounted equipment such as head tracking, eye tracking, and head mounted display components.
- A magnetic tracking device determines the position and orientation of a magnetic sensor (actually a reference frame attached to the sensor) with respect to a reference

frame fixed to the magnetic transmitter. For head tracking, the sensor with its imaginary attached reference frame is fastened to the pilot's head gear. If a similar device is being used to measure hand position (gesture control) the sensor is fastened to the pilot's hand or wrist.

- A head mounted eye tracker typically reports data with relation to a head fixed reference frame, but not necessarily with the same origin or orientation as the eye frame described above.
- A head mounted display may project an image on a virtual surface that moves with the head and is represented by yet another reference frame.

Note that the cockpit, instrument panel, and magnetic transmitter reference frames, as described above, may have different positions and orientations, but are all fixed to the airframe and do not move with respect to each other. Likewise, the helmet, magnetic sensor, and head mounted display reference frames all move together, and always maintain the same relationship to one another. The reference frames attached to the pilot's head gear do move with respect to those that are fixed to the airframe. The head and eye reference frames may sometimes move with respect to the helmet reference frame (helmet slippage), and the relation between these frames may differ somewhat from pilot to pilot.

Relationships between reference frames that are fixed with respect to each other are relatively straight forward, while relationships between frames that move with respect to one another are more complex. Reference frame definitions are chosen to facilitate computations. A particular computational problem may be solved using a subset of the reference frames described above or may use reference frame definitions that differ from these examples, but the principles remain the same.

E.3 MATHEMATICAL NOTATIONS AND CONVENTIONS

Cartesian coordinates have a "handedness" determined by the choice of positive direction for any axes with respect to the other two. By convention a right handed coordinate system is formed when the positive directions of the X, Y, and Z axes have the same relation as the index finger, middle finger, and thumb of the right hand, when held as shown in Figure E-2. A left handed coordinate system conforms to the same relation when the left hand is used. In this section, we will always assume right handed Cartesian coordinates. For the purpose of matrix notation subscripting, the X, Y, and Z axes are the 1st, 2nd, and 3d axes respectively.

A vector is often expressed as

$$R = xi + yj + zk. \tag{1}$$

where i, j, and k are unit vectors in the directions of the three coordinate axis of a reference frame. The same vector may also be expressed as column matrix

$$R^n = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

where the superscript n denotes the coordinate frame in which x, y, and z are specified. The vector component values (x, y, z) are sometimes referred to as the direction numbers for a line parallel to the vector. (2)

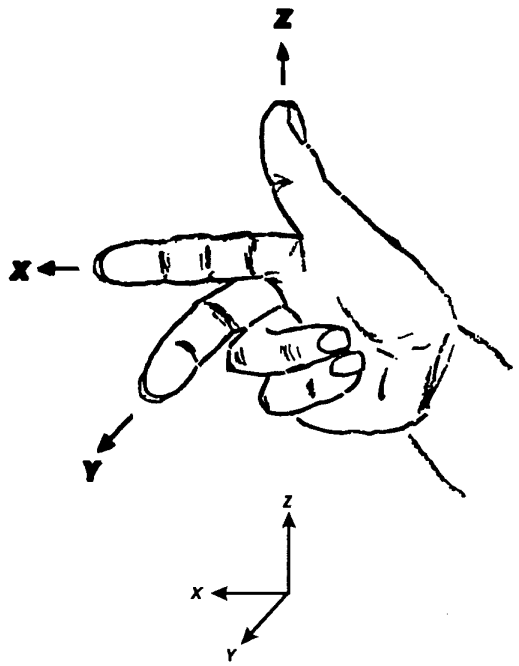


Figure E-2. Right handed Cartesian coordinates

Note that the vector has a direction and length, but its starting point is not specified. If a vector is assumed to start at the origin of the coordinate frame, then its component values are the coordinates of its end point, as shown in Figure E-3. Vectors can be added geometrically by placing them tip to tail, and mathematically by adding corresponding components (so long as all components are specified in the same reference frame).

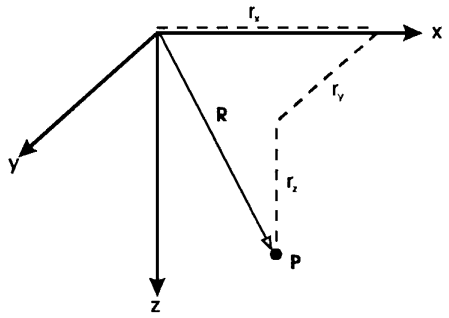


Figure E-3. Vector R and point P. The components of vector R (r_x, r_y, r_z) are also the coordinates of point P.

A plane can be specified in a given reference frame by the equation

$$ax + by + cz + d = 0 \tag{3}$$

where a, b, c, and d are constants, and x, y, and z are variables denoting the coordinates of any point on the plane. The vector formed by components (a, b, c) is perpendicular to the plane. Given any vector with components (a, b, c) and a point with coordinates (x_1, y_1, z_1), the plane perpendicular to

the vector and containing the point is given by equation (3) where

$$d = -(ax_1 + by_1 + cz_1). \tag{4}$$

A line in space can be specified in a given reference frame by the parametric equation

$$\begin{aligned} x &= x_0 + t(a) \\ y &= y_0 + t(b) \\ z &= z_0 + t(c) \end{aligned} \tag{5}$$

where (a, b, c) are the components of any vector parallel to the line and (x_0, y_0, z_0) are the coordinates of any point on the line. A point with coordinate values (x, y, z) calculated for any value of t must lie on the line. Note that the line has an absolute position and orientation in space, but has infinite length.

More complex curves and surfaces can also be represented by appropriate equations, and algebraic techniques can be used to solve for the intersections of lines and surfaces.

The location of one reference frame with respect to another is specified by the vector that connects the two origins. The rotation of one Cartesian coordinate frame with respect to another is most commonly specified by either Euler angles, a direction cosine matrix, or a quaternion.

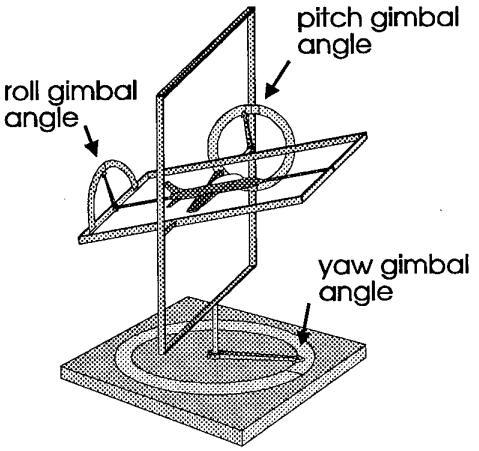


Figure E-4. Gimbals illustrating Euler angles

Euler angles, created by rotations about the three coordinate axes in a specified order, are illustrated by the conventional yaw (ψ), pitch (θ), and roll (ϕ) angle designations for aircraft attitude with respect to a navigational reference frame. Rotation order is critical and must be specified. In the general aircraft convention, the first rotation (ψ) is about the navigation frame z axis, producing modified axes x' and y'. The second rotation (θ) is about the y' axis, producing x'' and z''. The third rotation (ϕ) is about the x'' axis, finally resulting in axes parallel to the aircraft reference frame. This is illustrated by the gimbals shown in Figure E-4.

A problem with Euler angles is that they become indeterminate at certain attitudes. For example, when an aircraft is pointed straight up yaw (heading) and roll are undefined.

Rotation of frame a with respect to frame b can also be specified by a 3 by 3 matrix (C_b^a) for which each element c_{ij} is the cosine of the angle between i^{th} axis of frame a and the j^{th} axis of frame b. The cosine matrix defining rotation of b with respect to a is the inverse, of C_b^a. In the case of a direction cosine matrix, the inverse can be shown to be the same as the transpose.

$$C_a^b = (C_b^a)^{-1} = (C_b^a)^T \tag{6}$$

In terms of the Euler angles as defined above (rotation order from frame a to frame b is yaw, pitch, roll), the direction cosine matrix is

$$C_b^a = \begin{bmatrix} c\psi\, c\theta & s\psi\, c\theta & -s\theta \\ c\psi\, s\theta\, s\phi - s\psi\, c\phi & s\psi\, s\theta\, s\phi + c\psi\, c\phi & c\theta\, s\phi \\ c\psi\, s\theta\, c\phi + s\psi\, s\phi & s\psi\, s\theta\, c\phi - c\psi\, s\phi & c\theta\, c\phi \end{bmatrix} \tag{7}$$

(s ≡ sin; c ≡ cos)

Quaternions are a method for specifying rotation that has more recently become popular, especially for computer graphics manipulations. A quaternion consists of a set of 4 component values.

$$q = q_0 + q_1 i + q_2 j + q_3 k \tag{8}$$

The first component (q_0) is a scaler, and the last 3 (q_1, q_2, q_3) are the components of a vector. The vector direction represented by the last three quaternion components is the axis about which one frame would have to rotate to become parallel the other. The length of the vector is sin(θ/2), where θ is the angle through which the first frame would have to rotate about the specified axis to become parallel to the second. The first component (q_0) is a scalar quantity equal to cos(θ/2).

Another notation sometimes used for quaternion rotation specification is $e^{\frac{1}{2}\theta u}$, where u is a unit vector specifying a rotation axis and θ is the rotation angle. This is a compact notation arrived at by analogy with the mathematics of complex numbers, and *does not* imply the mathematical properties of the quantity e raised to a power.

A direction cosine matrix for rotation of frame a relative to frame b can be generated from a quaternion q as follows

$$C_b^a = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_1) \\ 2(q_3q_0 + q_1q_2) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_1q_0 + q_3q_2) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{9}$$

If the relative orientation of two reference frames are known, vector components specified in one frame can be transformed to equivalent components in the other frame by multiplying the vector column matrix by the direction cosine matrix.

$$R^a = C_b^a R^b \tag{10}$$

If the relative position of the two frames is also known, the coordinates of a point known in one frame can also be easily transformed to the other by a combination of vector

transformation and addition. Referring to Figure E-5, we assume that R^a_{ap}, R^a_{ab}, and C^b_a are known.

$$\begin{aligned} R^b_{ap} &= C^b_a R^a_{ap} \\ R^b_{ab} &= C^b_a R^a_{ab} \\ R^b_{bp} &= R^b_{ap} - R^b_{ab} \end{aligned} \tag{11}$$

The coordinates of point P in frame *b* are the components of R^b_{bp}.

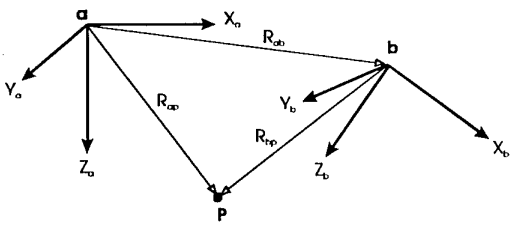


Figure E-5. Relation between a point P and two different coordinate frames (a and b)

A more thorough treatment of reference frames and coordinate transformations can be found in [E-1].

E.4. EXAMPLE REFERENCE FRAME RELATIONSHIP PROBLEM

Reference frame relationships can be illustrated by the problem of determining eye point of gaze on a display using information from a head mounted eye tracker and a magnetic head tracker.

E.4.1 Simplified problem

The problem is illustrated by Figure E-6. To simplify the example, we will make several assumptions

1. The magnetic transmitter is fixed to the airframe
2. The position of the display reference frame origin and its orientation are known in transmitter frame coordinates.
3. The magnetic sensor is fastened directly to the pilot's head (so that we need not be concerned with head/helmet slippage).
4. The vector from the sensor to the eye is known in the sensor coordinate frame.
5. Orientation of the eye reference frame with respect to the sensor frame is known, and the eye tracker reports line of gaze by specifying its direction with respect to the eye reference frame.
6. The head tracker reports the orientation and position of the sensor frame with respect to the transmitter frame.

Some of these assumptions, especially numbers 3 and 5, are unrealistic, and the implications are discussed later on. The problem is to determine point of gaze on the display in display reference frame coordinates, since it is in this frame (or a closely related pixel coordinate frame) that we know the position of display elements.

The problem can be solved by using head tracker information to develop the equation, in transmitter coordinates, for the line along which gaze falls. It is then possible to solve for the intersection of a line (line of gaze) and a plane (display surface) in transmitter coordinates, yielding point of gaze expressed in the transmitter coordinate frame. Vector arithmetic can then be used to find the vector from the display frame origin to the point of gaze, also expressed in transmitter coordinates; and finally a coordinate transformation of this vector to the display coordinate frame yields the desired result. Note that these calculations must be done at whatever frequency point of gaze is to be updated. As with reference frame definitions, the details of the computation sequence may vary from the above description, but the principles remain the same.

E.4.2 Mathematical treatment of simplified problem

Referring to Figure E-6, the problem is to find the point of gaze in display coordinates (Rdp^{dsply}). The assumptions listed in the preceding section translate to the following list of known mathematical quantities.

- **Known vectors:** $Rtd^{trnsmtr}$, Rse^{sensor}, $Rts^{trnsmtr}$, direction (but not length) for Rep^{eye}
- **Known direction cosine matrices:** $C^{trnsmtr}_{sensor}$, C^{sensor}_{eye}, $C^{trnsmtr}_{dsply}$

The vector from the transmitter to the origin of the display reference frame (Rtd), and the vector from the sensor to the eye (Rse), assumed to be known in the transmitter and sensor coordinate frames respectively, are determined in advance. $Rtd^{trnsmtr}$ and the relative attitude of the transmitter and display coordinate frames ($C^{trnsmtr}_{dsply}$) are constant so long as equipment remains mounted to the same place in the cockpit. Rse^{sensor} and C^{sensor}_{eye} remain constant once the magnetic sensor is fastened to the pilot's head.

The plane of the display is coincident with the *yz* plane of the display coordinate frame, and its equation in the transmitter frame can be quickly calculated in advance using equations (3) and (4) as shown below by equations (12) through (14). A unit vector in the direction of the display frame z axis is known to be orthogonal to the plane and can be transformed to transmitter coordinates by

$$Ud^{trnsmtr} = \begin{bmatrix} ud_1 \\ ud_2 \\ ud_3 \end{bmatrix} = C^{trnsmtr}_{dsply} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{12}$$

We also know that the origin of the display reference frame is a point on the plane and has transmitter frame coordinates given by

$$Rtd^{dsply} = \begin{bmatrix} rtd_1 \\ rtd_2 \\ rtd_3 \end{bmatrix}. \tag{13}$$

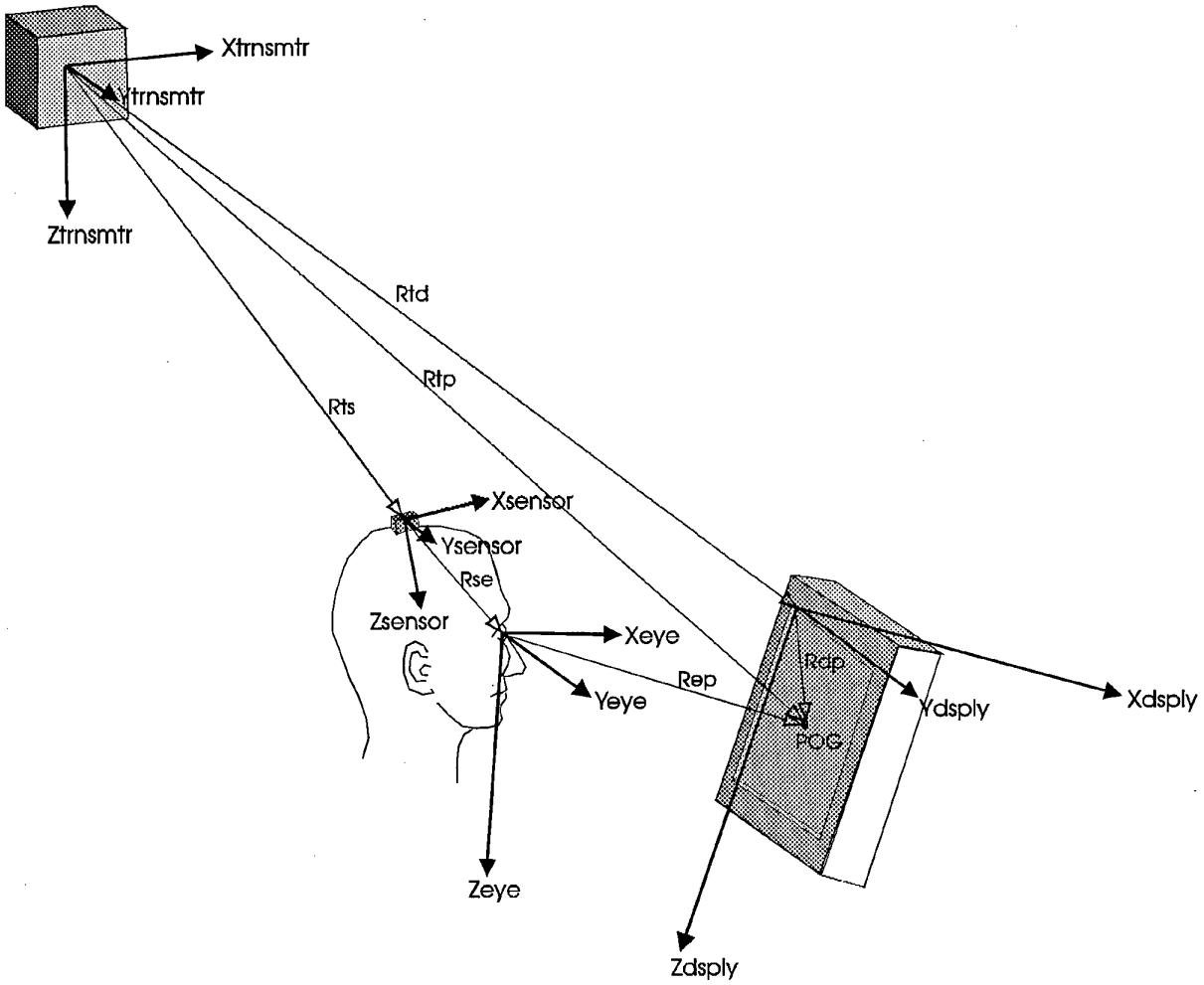


Figure E-6. Diagram showing reference frames and vectors that define the relationships between a magnetic transmitter, a head mounted magnetic sensor, a person's eye, and the person's point of gaze on a display panel. Coordinate frame axes are labelled with upper case X, Y, and Z followed, in lower case, by the reference frame name. Vectors are labelled with upper case R followed, in lower case, by the vector name (usually the first letter of the name of each end point). Point of gaze on the display is labelled "POG".

An equation for the display plane, in transmitter coordinates, is then

$$ud_1 x + ud_2 y + ud_3 z - (ud_1 rtd_1 + ud_2 rtd_2 + ud_3 rtd_3) = 0 \quad (14)$$

The remaining calculations must be repeated every time point of gaze is updated. The vector from the transmitter to the sensor ($Rts^{trnsmtr}$), and orientation of the sensor ($C_{sensor}^{trnsmtr}$), continually change as the pilot's head moves about, and these are the quantities reported by the magnetic head tracker. We will assume that line of gaze direction, as reported by the eye tracker, is available as a unit vector (Uep^{eye}) in the eye coordinate frame. First we translate all pertinent vectors to transmitter coordinates.

$$\begin{aligned} Rse^{trnsmtr} &= C_{sensor}^{trnsmtr} Rse^{sensor} \\ Uep^{trnsmtr} &= \begin{bmatrix} uep_1 \\ uep_2 \\ uep_3 \end{bmatrix} = C_{sensor}^{trnsmtr} C_{eye}^{sensor} Uep^{eye} \end{aligned} \quad (15)$$

Location of the eye reference frame in transmitter coordinates is

$$Rte^{trnsmtr} = \begin{bmatrix} rte_1 \\ rte_2 \\ rte_3 \end{bmatrix} = Rts^{trnsmtr} + Rse^{trnsmtr} \quad (16)$$

From (5), the parametric equation for the line of gaze, in transmitter coordinates is

$$\begin{aligned} x &= rte_1 + t(uep_1) \\ y &= rte_2 + t(uep_2) \\ z &= rte_3 + t(uep_3) \end{aligned} \quad . \qquad (17)$$

Point of gaze is on the display plane, and is also on the line of gaze. By substituting (17) into (14), solving for t, and then substituting t back into (17), we get the intersection of the line (17) and plane (14). The resulting coordinates are the components of $Rtp^{trnsmtr}$. Finally, by vector addition and a coordinate transform

$$\begin{aligned} Rdp^{trnsmtr} &= Rtp^{trnsmtr} - Rtd^{trnsmtr} \\ Rdp^{dsply} &= C^{dsply}_{trnsmtr} Rdp^{trnsmtr} \end{aligned} \qquad (18)$$

E.4.3 Additional complexity

In the real world situation, a magnetic sensor must be fixed to the pilot's head gear rather than directly to his head. Because of anatomical variations, the vector connecting the sensor and the eye (in sensor coordinates) will vary from pilot to pilot; and because of head/helmet slippage the vector will not be perfectly constant even for a given pilot.

Typical eye trackers do not directly provide line of gaze direction with respect to a precisely known eye reference frame as shown in Figure E-6. Generally a calibration procedure maps the raw data (e.g. pupil to corneal reflection vector) to gaze points on a head mounted display projection or a physical surface in the cockpit. That information must be converted to line of gaze with respect to an eye coordinate frame or some other related reference frame to enable the type of computations described in the previous sections.

The additional complexity is typically handled by making additional measurements, or by manipulating the computations and the placement of components so as to minimise result sensitivity to the uncertain variables.

Detailed implementation examples are beyond the scope of this document. There are many equivalent formulations for solving similar reference frame relationship problems, and the example presented is intended to be illustrative.

E.5 REFERENCE

E-1. Wrigley, W., Hollister, W. M., and Denhard, W. G., "Gyroscopic Theory, Design, and Instrumentation", Cambridge, The M.I.T. Press, 1969.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-7 AC/323(HFM)TP/3	ISBN 92-837-1009-6	UNCLASSIFIED/ UNLIMITED

5. Originator Research and Technology Organization
North Atlantic Treaty Organization
BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Alternative Control Technologies

7. Presented at/sponsored by
The RTO Human Factors and Medicine Panel (HFM).

8. Author(s)/Editor(s)	9. Date
Multiple	December 1998
10. Author's/Editor's Address	**11. Pages**
Multiple	148

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Human factors engineering	Motion studies
Artificial intelligence	Eye movements
Man machine systems	Man computer interface
Control equipment	Computerized simulation
Speech recognition	Voice communication
Cockpits	Head (anatomy)
Adaptive systems	Eye (anatomy)
Automatic control	Tracking (position)
Integrated systems	Electrophysiologic recording
Pilots (personnel)	Fighter aircraft

14. Abstract

In January 1996, the Working Group 25 of the former AGARD Aerospace Medical Panel began to evaluate the potential of alternative (new and emerging) control technologies for future aerospace systems. The present report summarizes the findings of this group. Through different chapters, the various human factors issues related to the introduction of alternative control technologies into military cockpits are reviewed, in conjunction with more technical considerations of the state of the art of the enabling technologies. Cockpit integration issues are emphasized in regard to both human factors and engineering constraints. Several specific applications of these new technologies in the aerospace environment are considered. Challenges for further developments are identified and recommendations issued. Globally, based upon operational condiderations and currently recognized limitations of the HOTAS concept, the conclusion is that Alternative Control Technology should now be progressively introduced into the cockpit, as a function of degree of maturity of the various techniques.

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Directeur - Gestion de l'information
(Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
RTO Coordination Office
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone**.

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Director Research & Development
Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
RTO Coordination Office
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-837-1009-6